

# СЕТЕВОЙ АНАЛИЗ В ГУМАНИТАРНЫХ ОБЛАСТЯХ

---



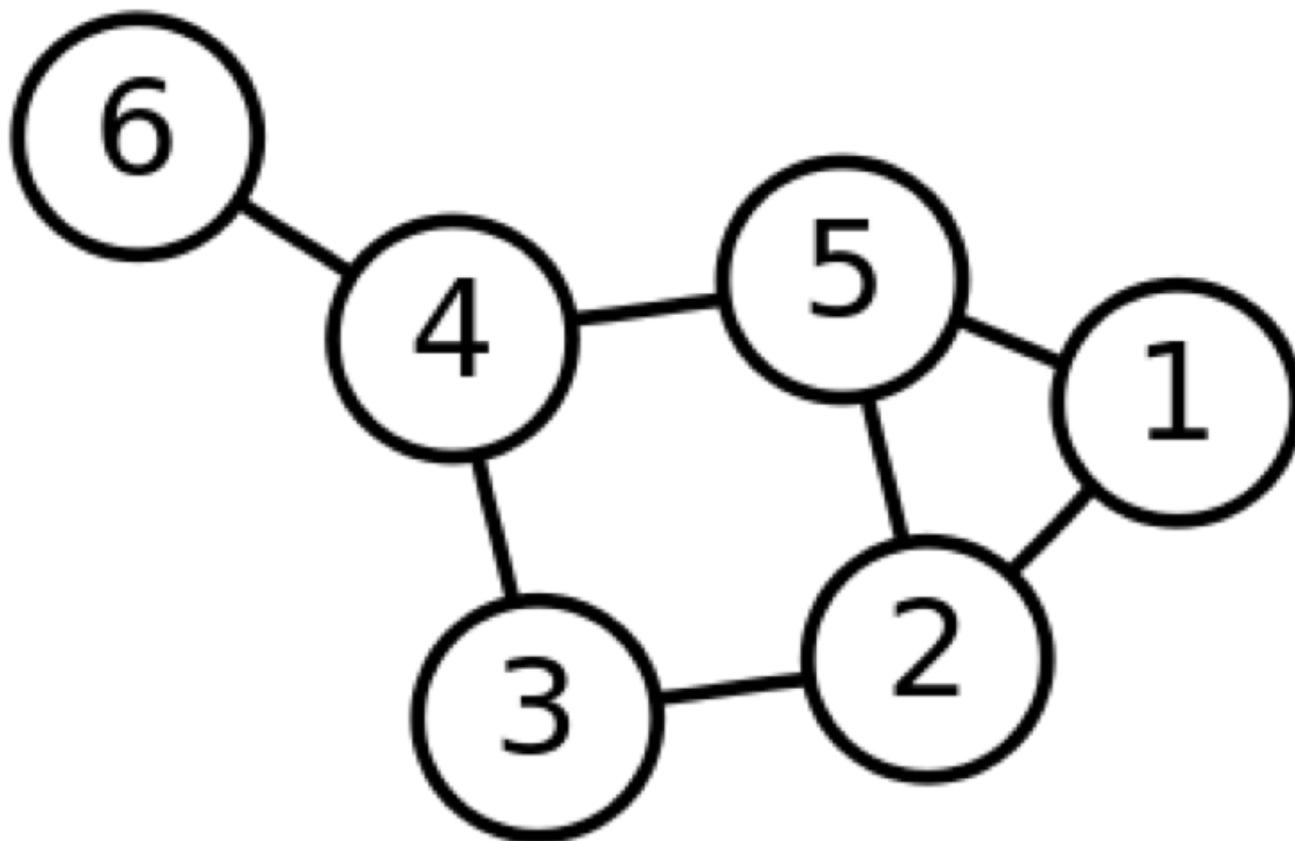
Современные методы в  
гуманитарных науках

3 марта 2018

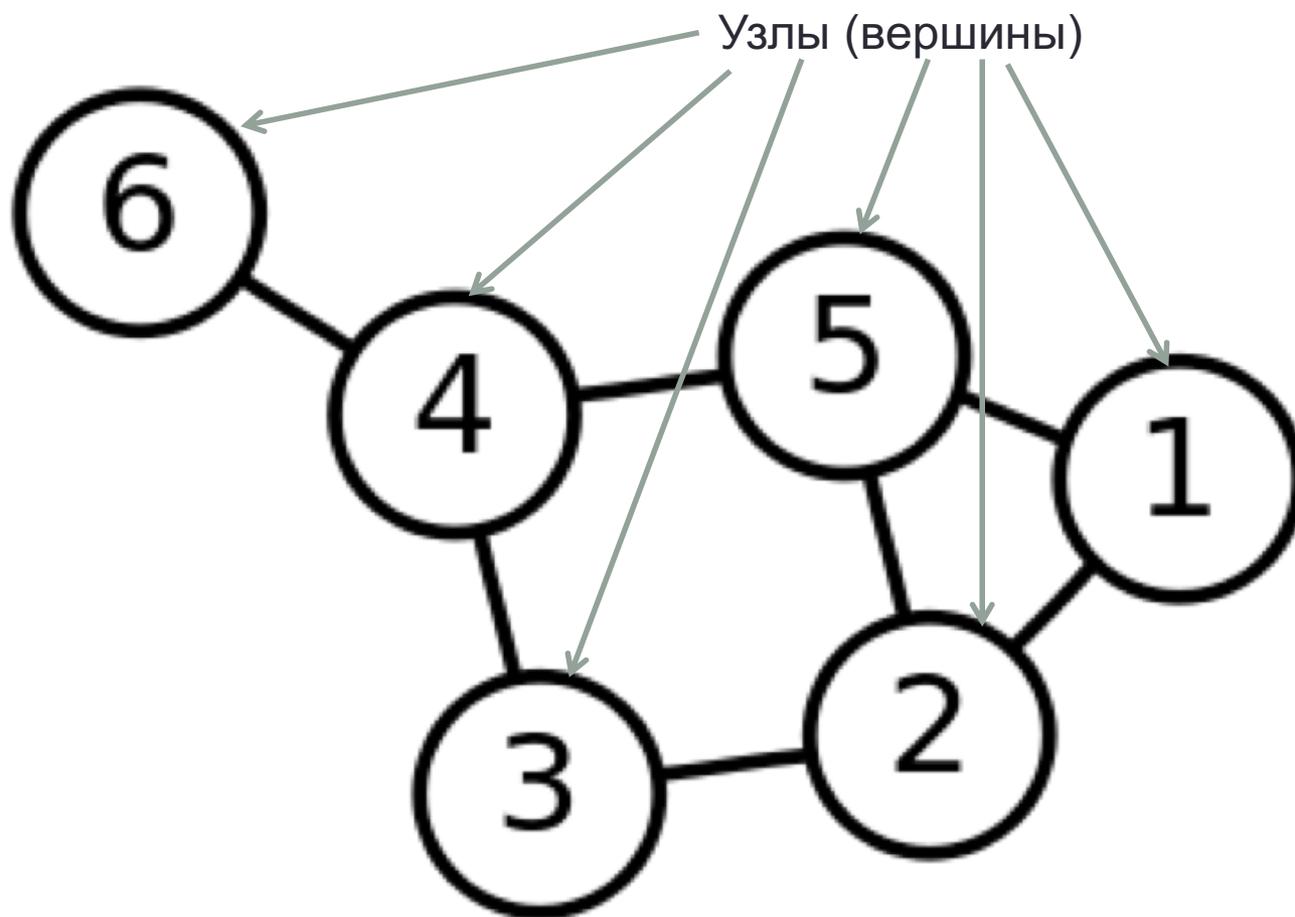
# О чем мы будем говорить

- Что за сети? Что за анализ?
- Те самые «шесть рукопожатий»: сети (графы) в реальном мире
- С кем дружил Фрэнсис Бэкон и кому платили короли: сети в гуманитарных науках
- Пушкин на посылках и «коммунистическая» пьеса: что можно увидеть в вымышленных сетях?
- Анализируем и визуализируем сети с помощью Gephi (практика)

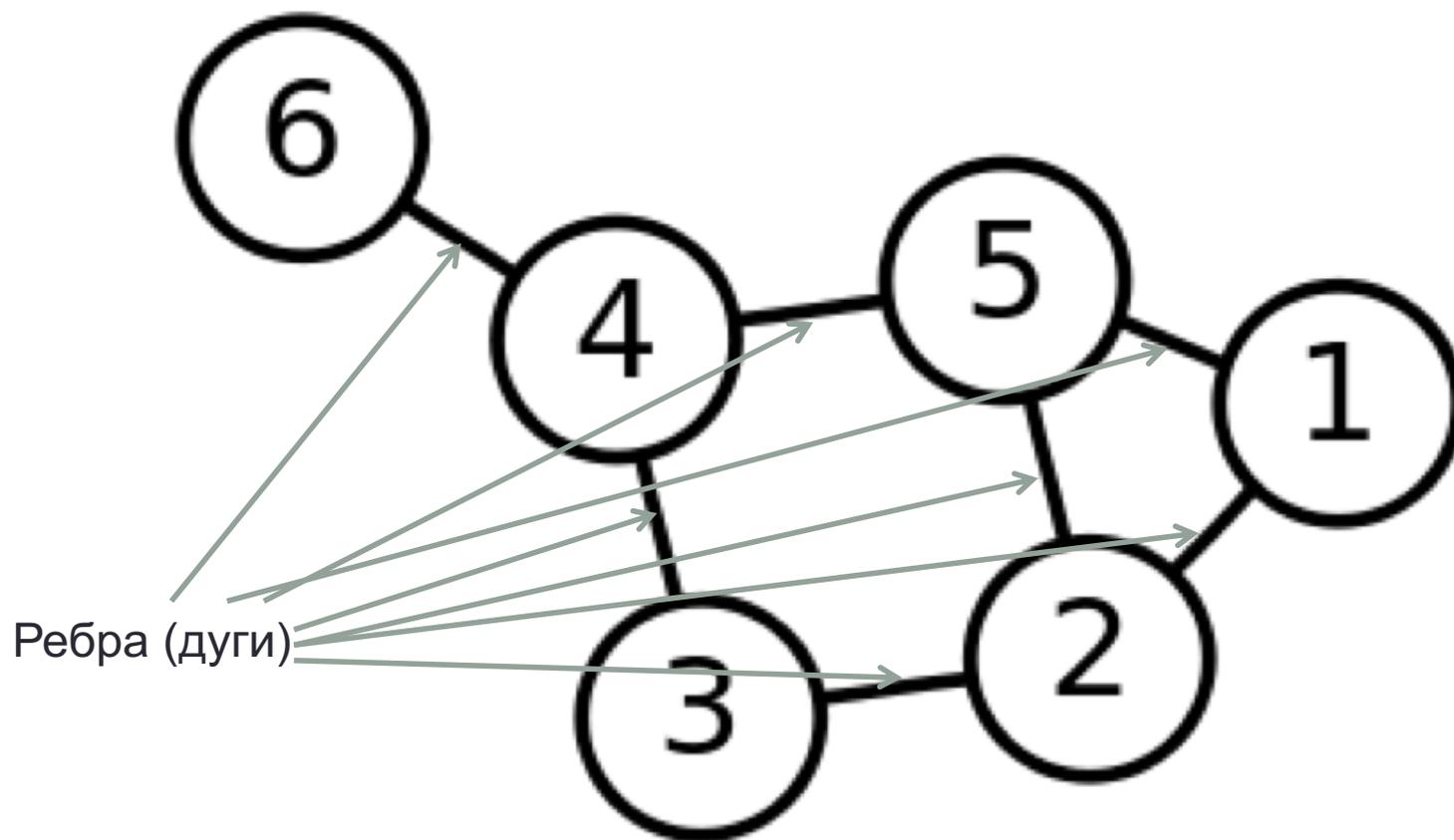
# Сеть (граф)



# Сеть (граф)

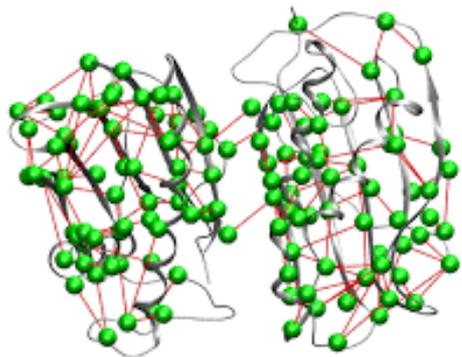
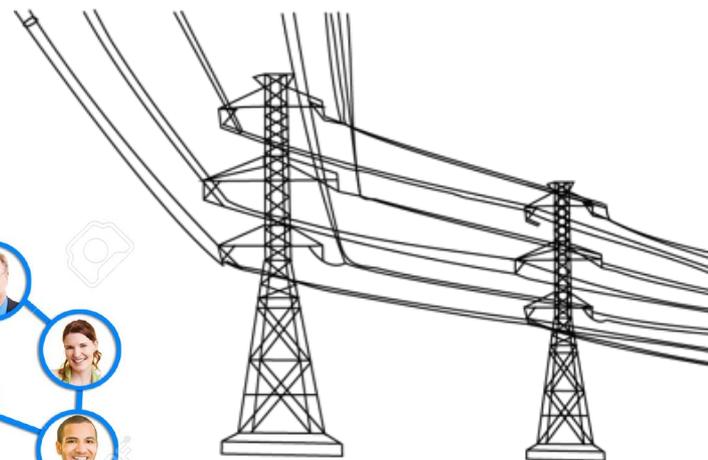
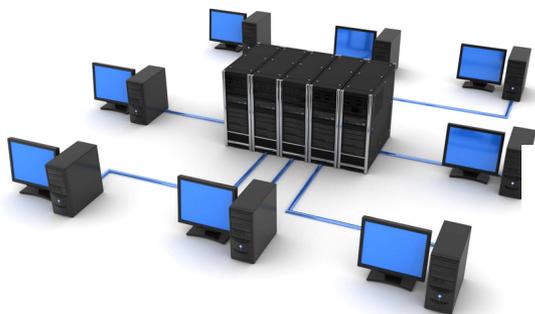


# Сеть (граф)

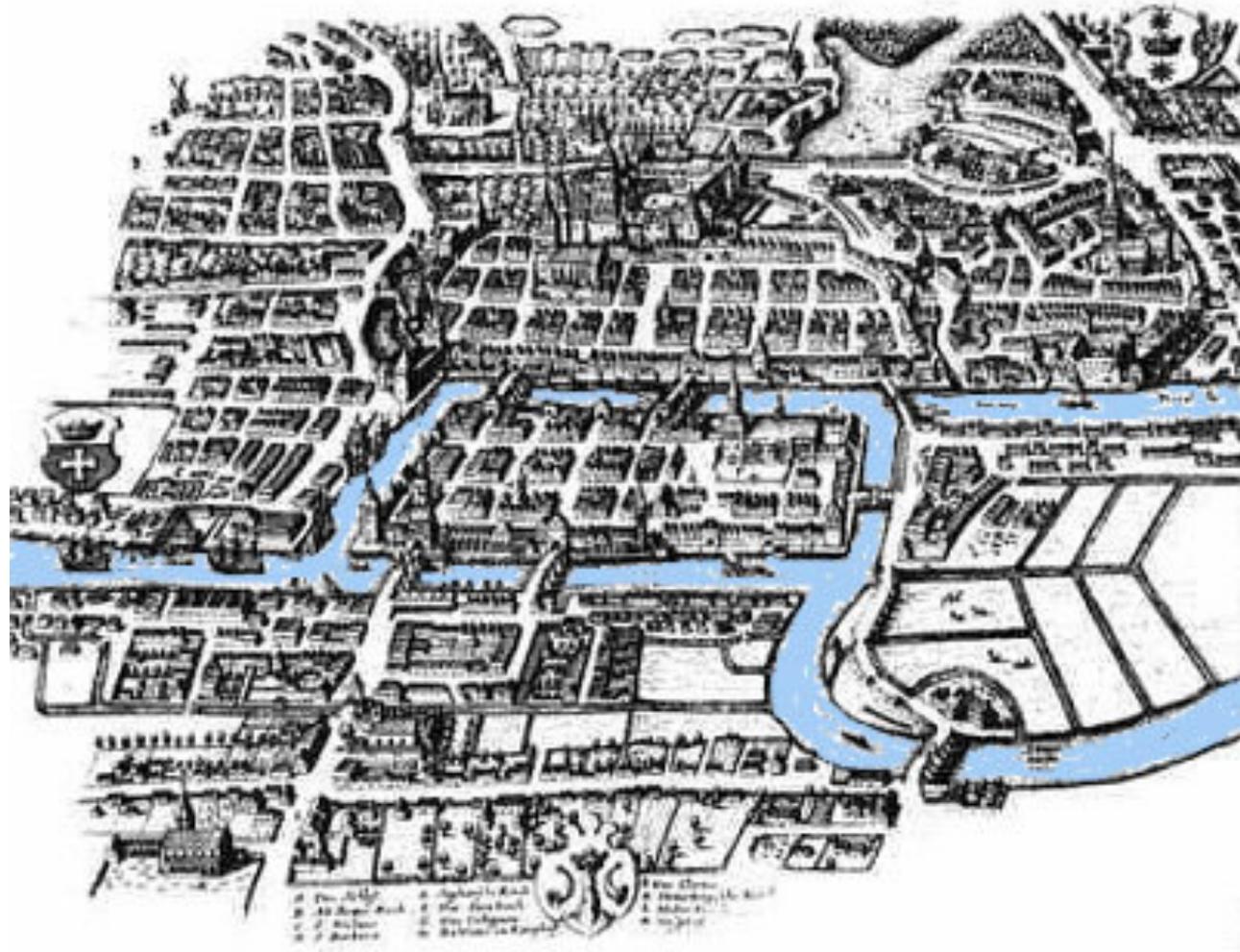


Что можно представить  
в таком виде?

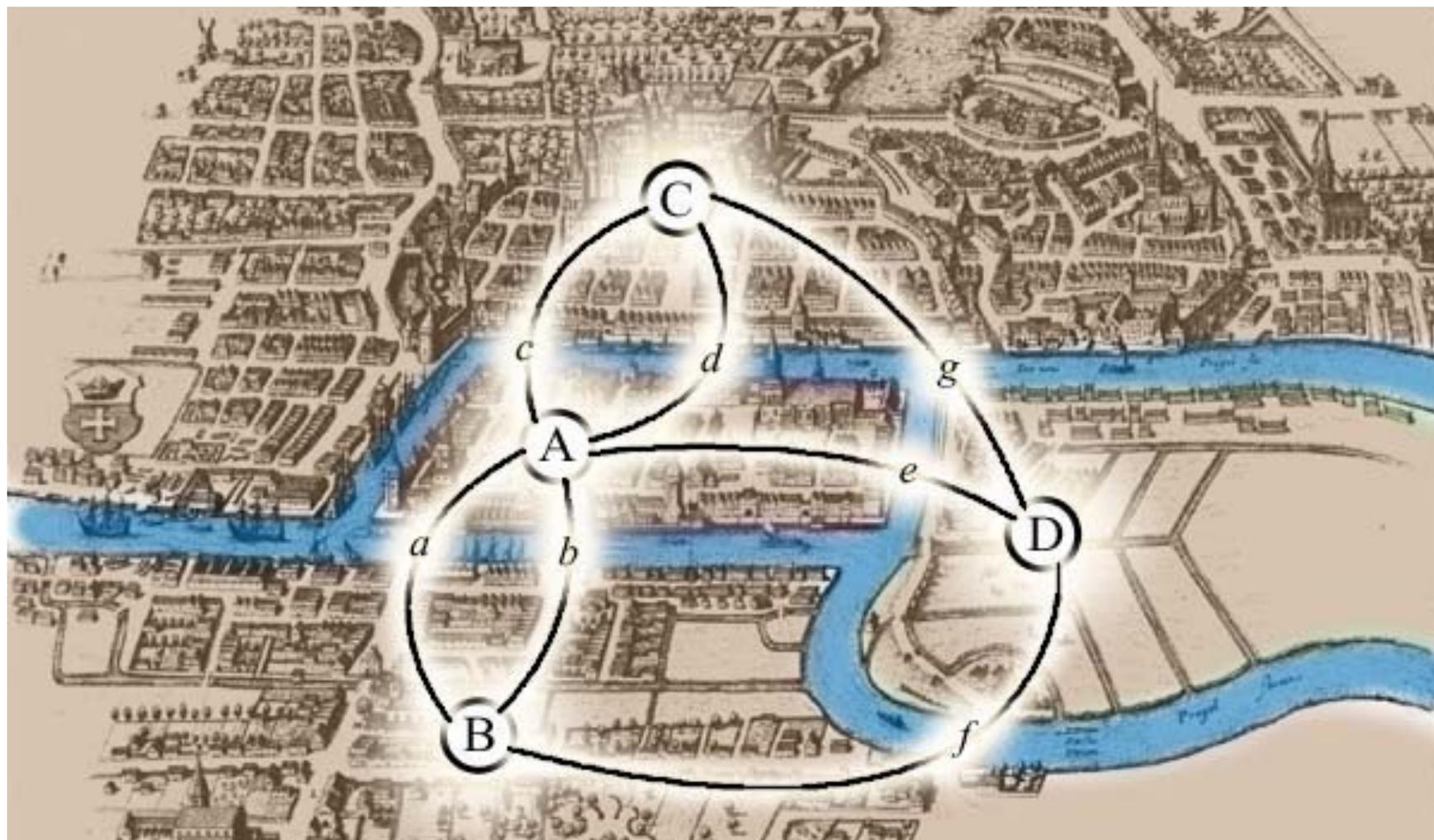
# Практически все



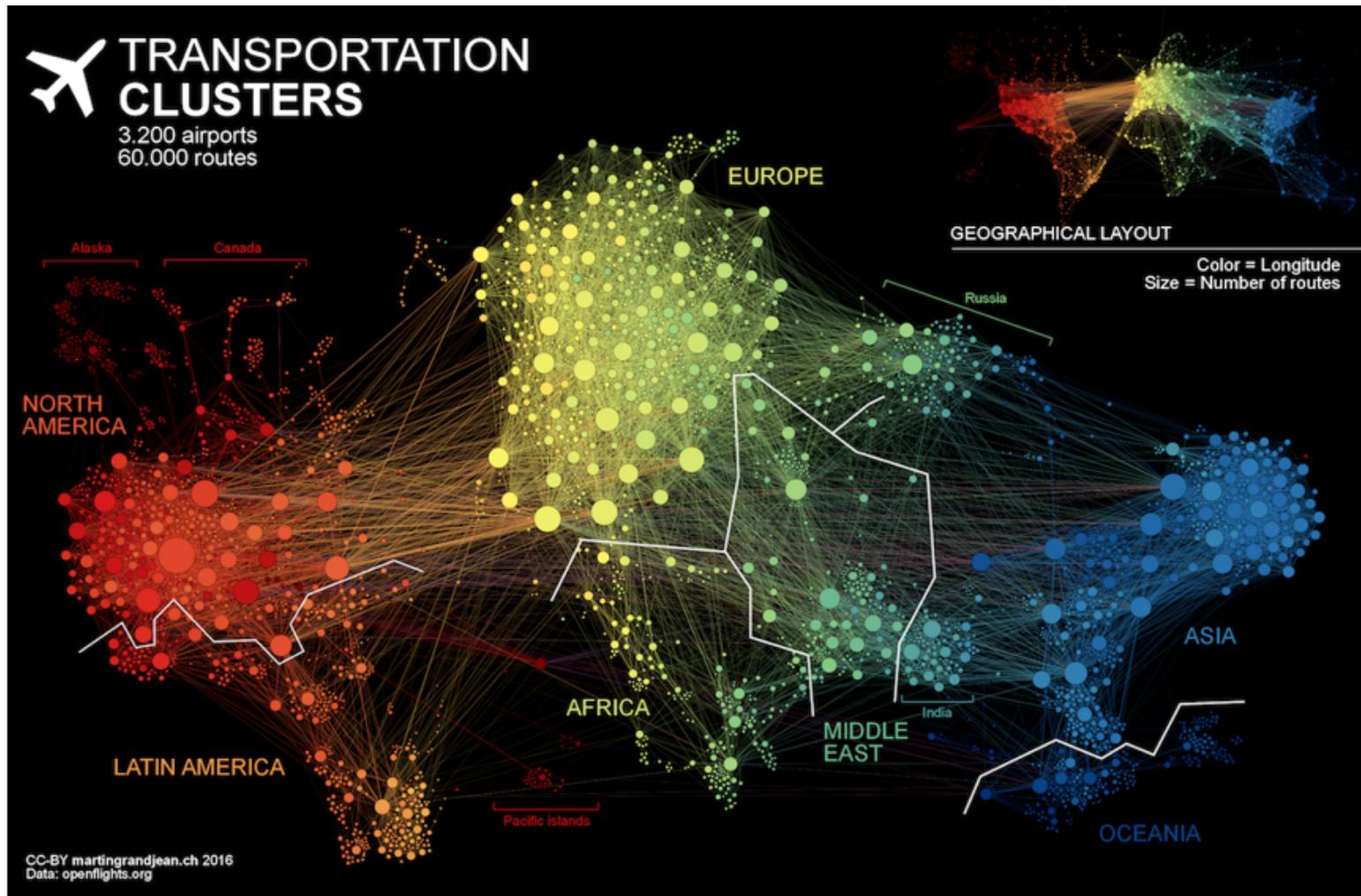
# Мосты (откуда есть пошли графы)



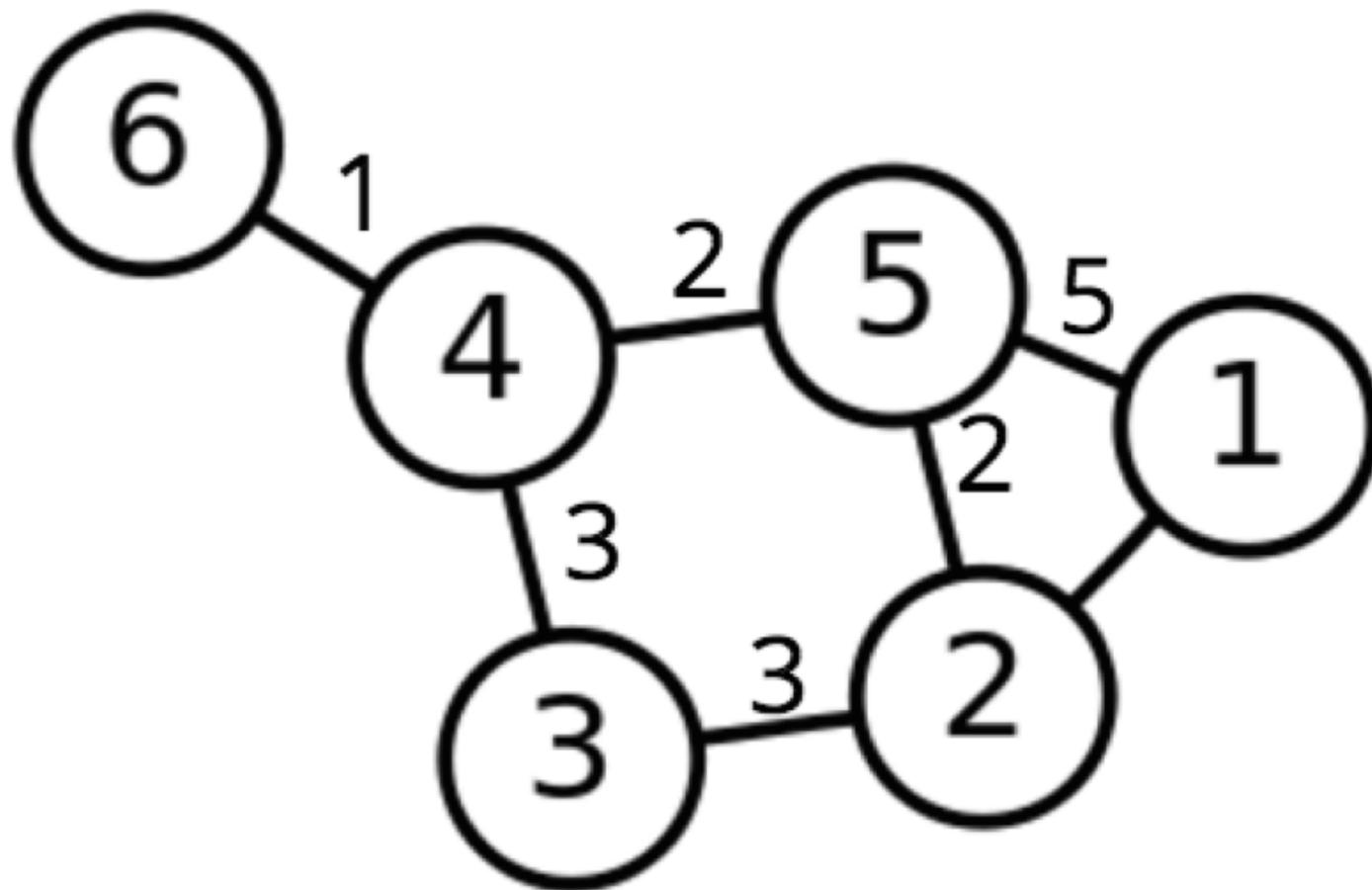
# Задача о 7 мостах Кёнигсберга



# Транспортные сети

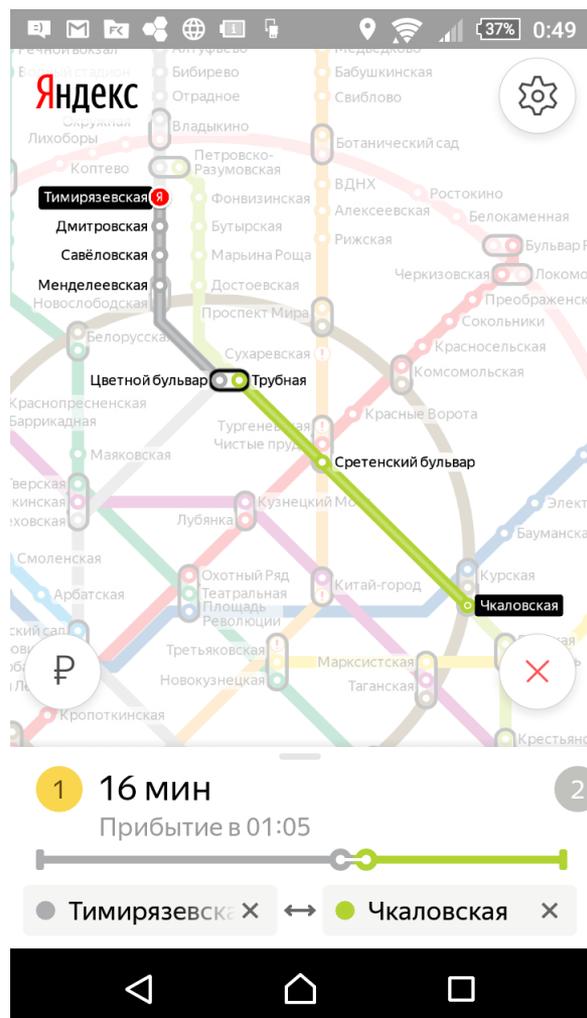


# Взвешенный граф





# Еще один взвешенный граф:

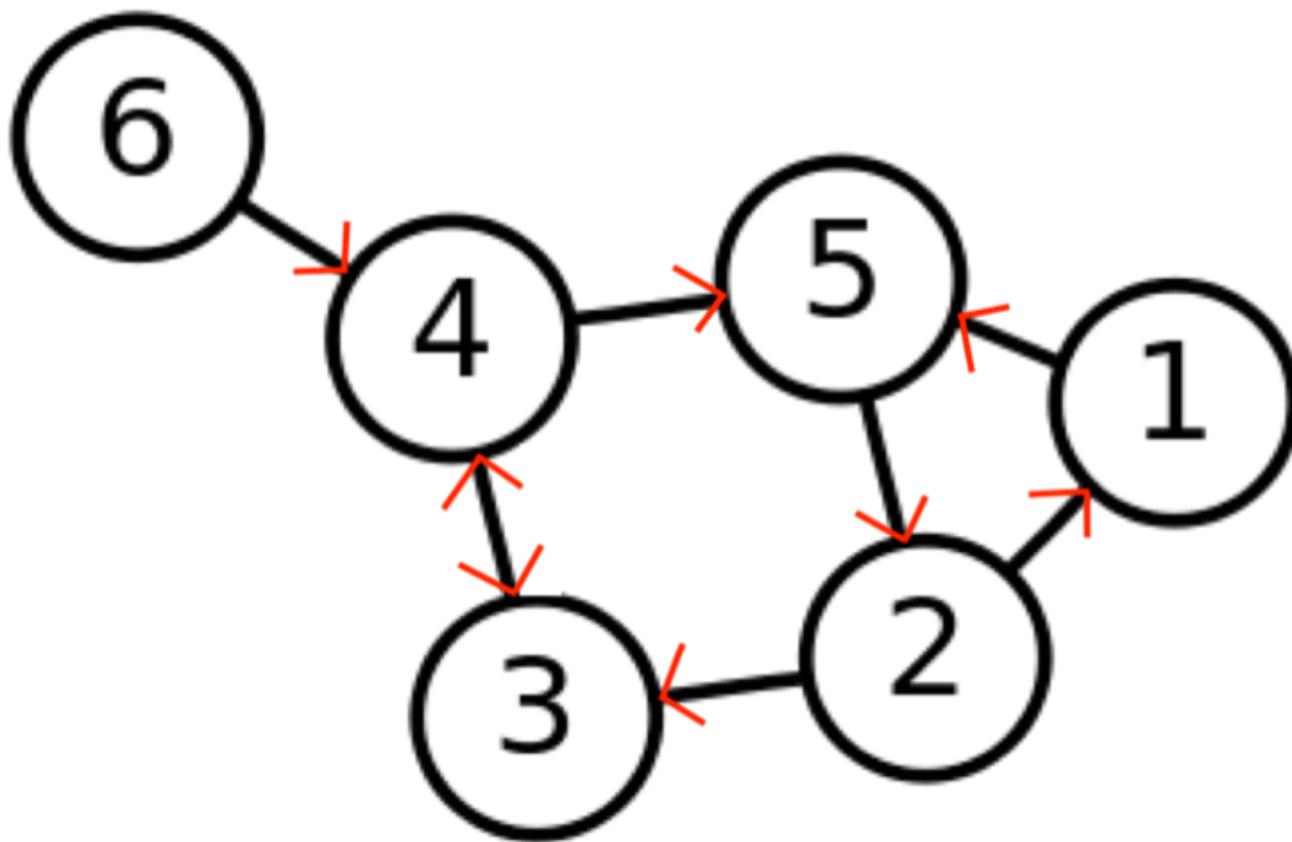


# Интернет



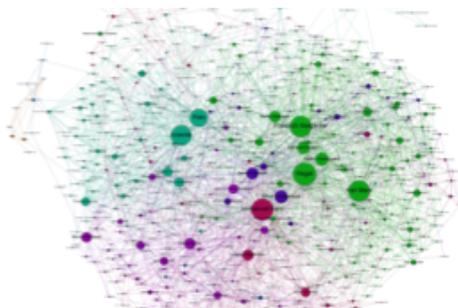


# Направленный граф



# Тизер на следующее занятие

- Такие графы



**Extracting Infobox Data**

```
http://en.wikipedia.org/wiki/Calgary
```

```
<http://dbpedia.org/resource/Calgary>  
dbpedia:native_name "Calgary" ;  
dbpedia:altitude "1048" ;  
dbpedia:population_city "988193" ;  
dbpedia:population_metro "1079310" ;  
mayor_name  
  dbpedia:Dave_Bronconnier ;  
governing_body  
  dbpedia:Calgary_City_Council ;  
...
```

■ Altogether 9,100,000 RDF triples  
extracted from 754,000 infoboxes

Christian Bizer et al.: DBpedia – Querying Wikipedia Like a Database (May 11, 2007)

Calgary	
	
Government	
- Mayor	Dave Bronconnier <small>(see mayor)</small>
- Governing body	Calgary City Council
- Manager	Owen A. Tobert
Area <sup>[?]</sup>	
- City	726.50 km² (280.5 sq mi)
- Metro	5,107.43 km² (1,972 sq mi)
Elevation	1,048 m (3,438.3 ft)
Population (2006) <sup>[?]</sup>	
- City	988,193
- Density	1,360.2/km² (3,522.9/sq mi)
- Metro	1,079,310
- Population rank	3rd
- Metro rank	5th

несложно доставать из Википедии самостоятельно с помощью



# ...и, конечно, социальные сети!

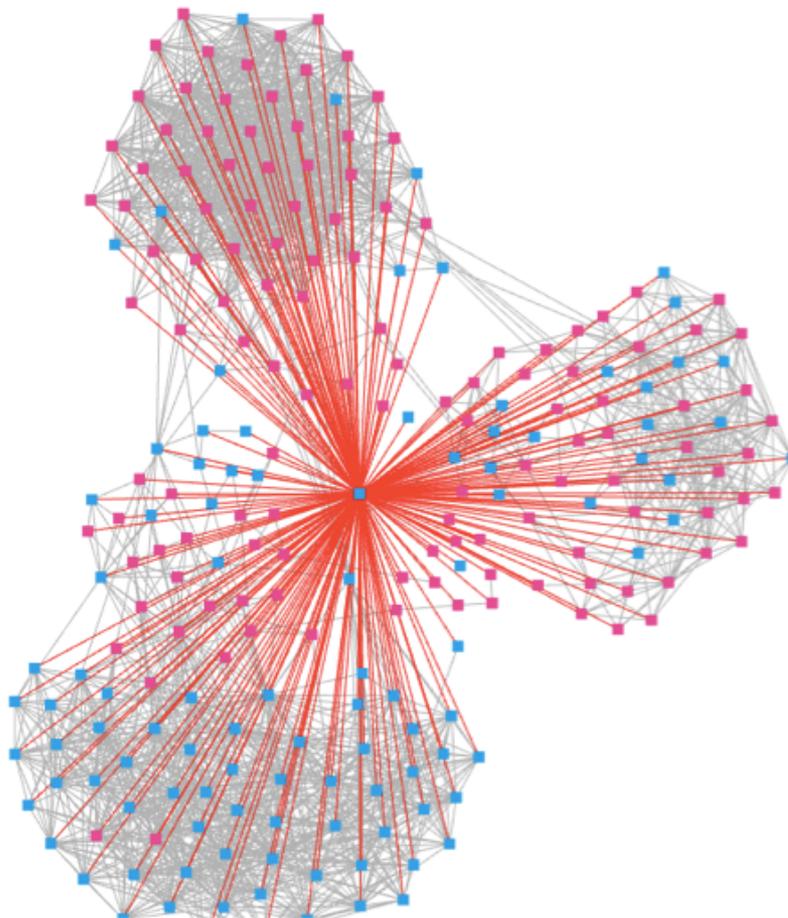
Показана 1 социальная сеть



+ Добавить сеть пользователя...

Сохранить картинкой...

234 людей; 2191 связь

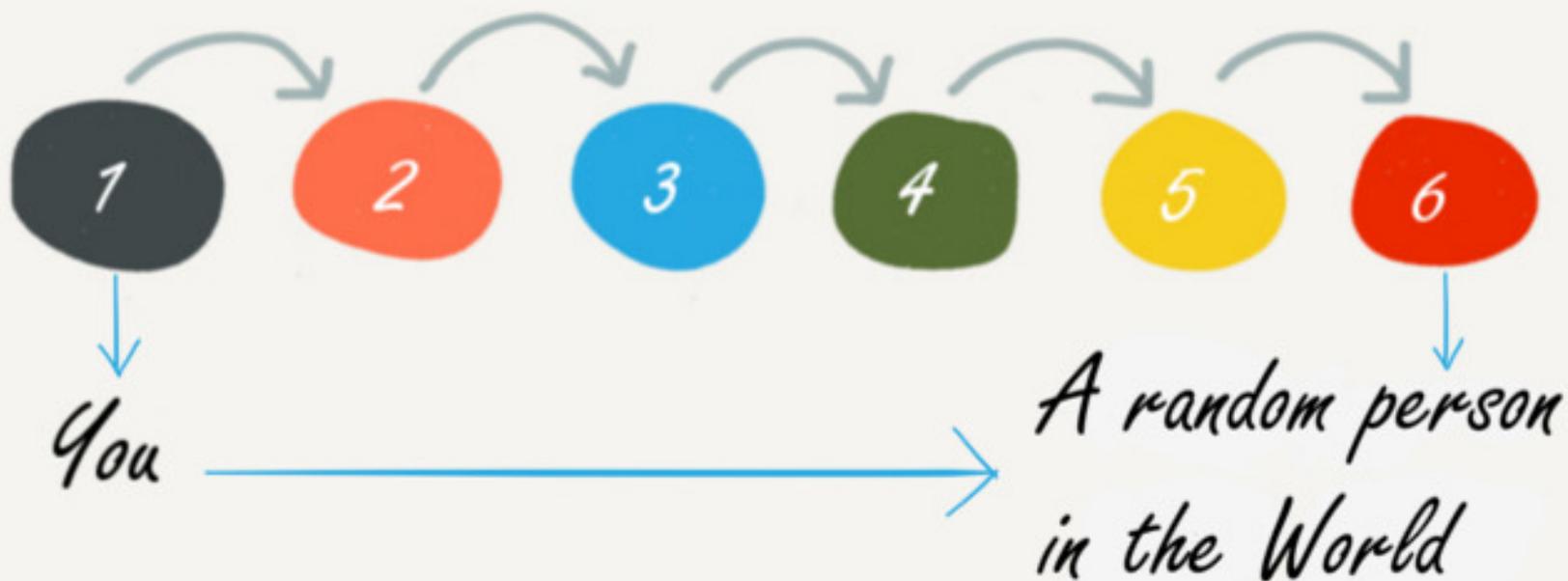


# Анализ социальных сетей (немного скучной истории)

- В современном виде появился в 1930-е годы в социальных науках (Курт Левин, Якоб Морено)
- Манчестерская школа антропологии в 1950-е (Макс Глакман и коллеги)
- «Гарвардский прорыв» 1960-х: бум социологии + компьютеры + структурализм
- Второй бум в 1990-е – 2000-е (интернет, вики-технологии, социальные сети)

# Шесть рукопожатий

## SIX DEGREES OF SEPARATION



# Шесть рукопожатий



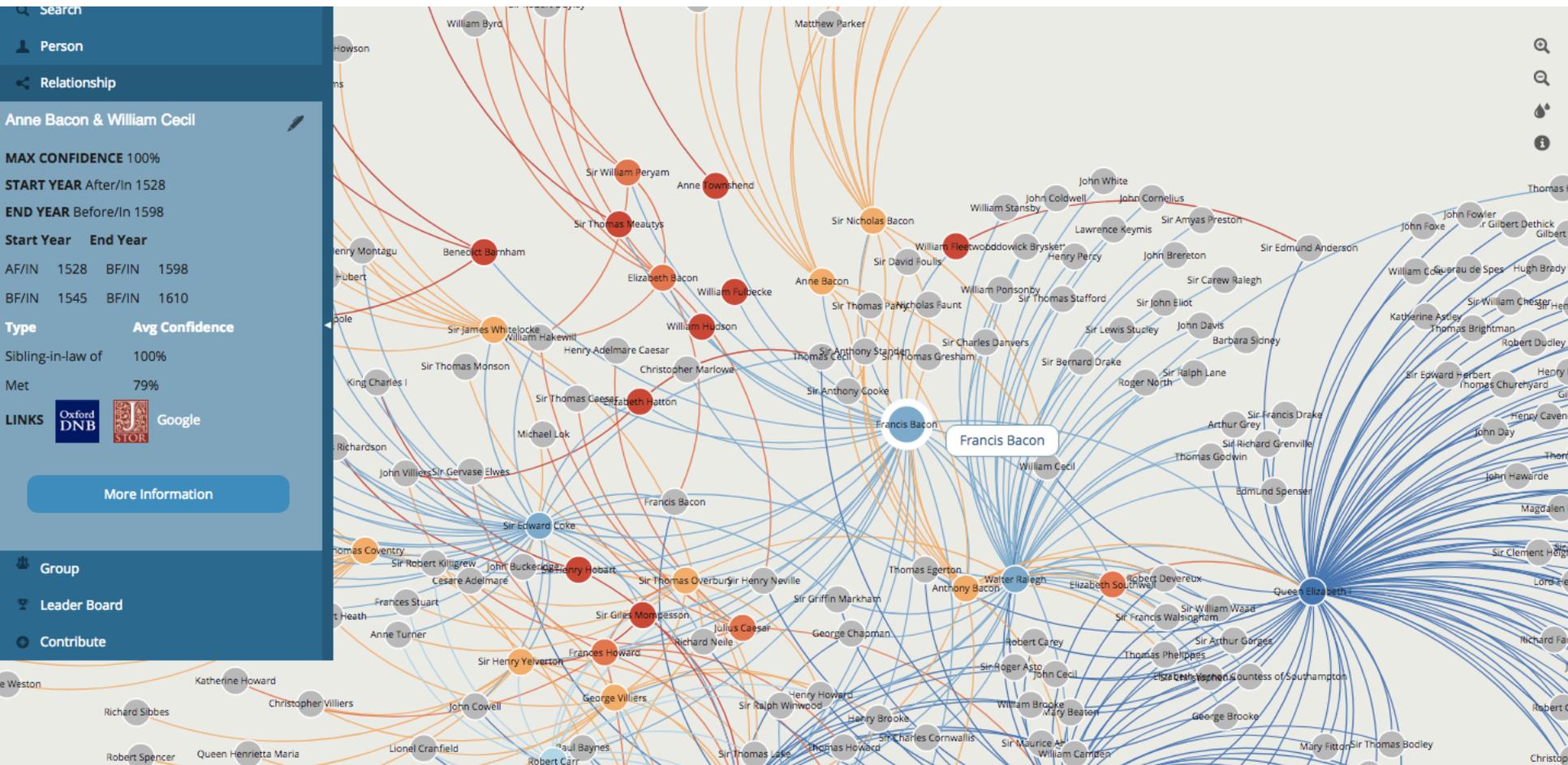
# СЕТЕВОЙ АНАЛИЗ В ГУМАНИТАРНЫХ НАУКАХ

---

# Кто кого знал в Британии 500 лет назад?

- Проект Six Degrees of Francis Bacon (шесть рукопожатий Фрэнсиса Бэкона)
- Более 13.000 человек, более 200.000 связей
- Извлечено из Oxford Dictionary of National Biography
- <http://www.sixdegreesoffrancisbacon.com/>

# Кто кого знал в Британии 500 лет назад?



# People of Medieval Scotland

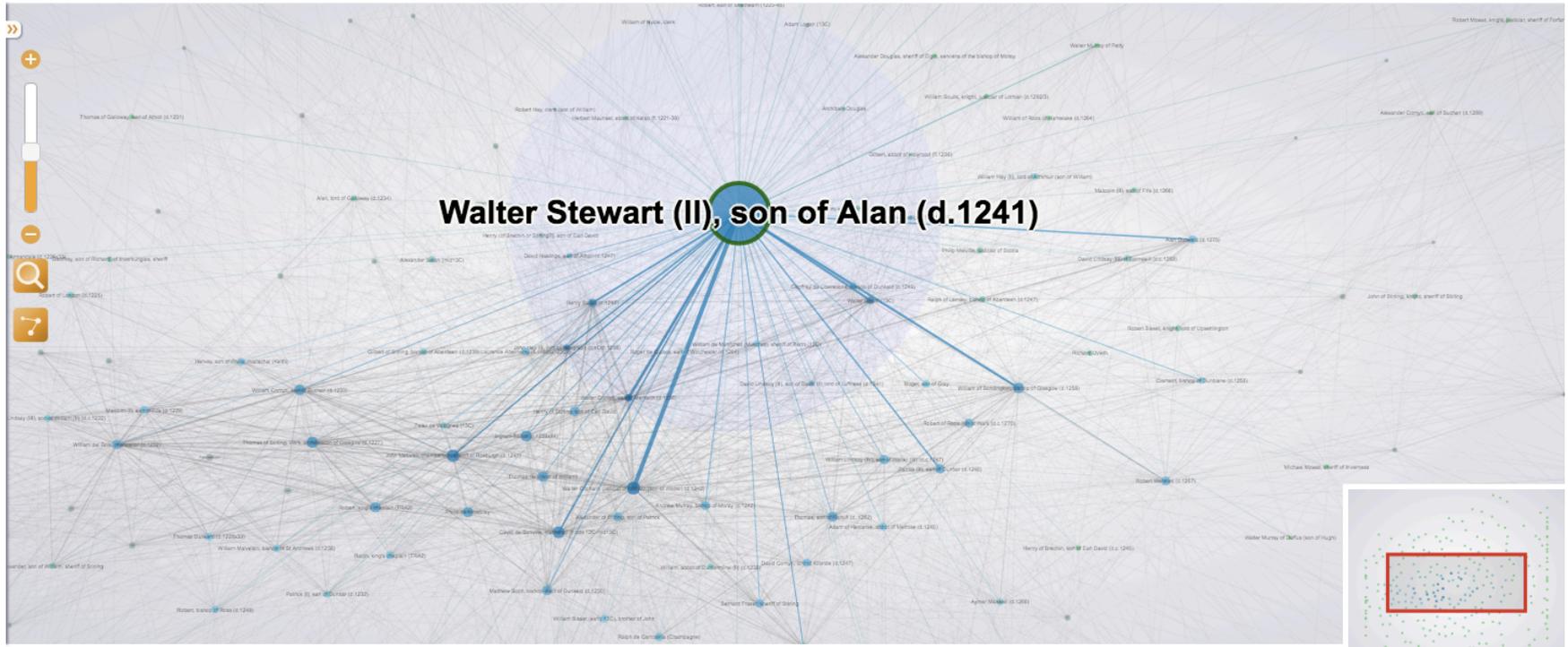
PEOPLE OF MEDIEVAL SCOTLAND  
1093-1314

Home About Database Information Schools News Contact Social Network Analysis Help

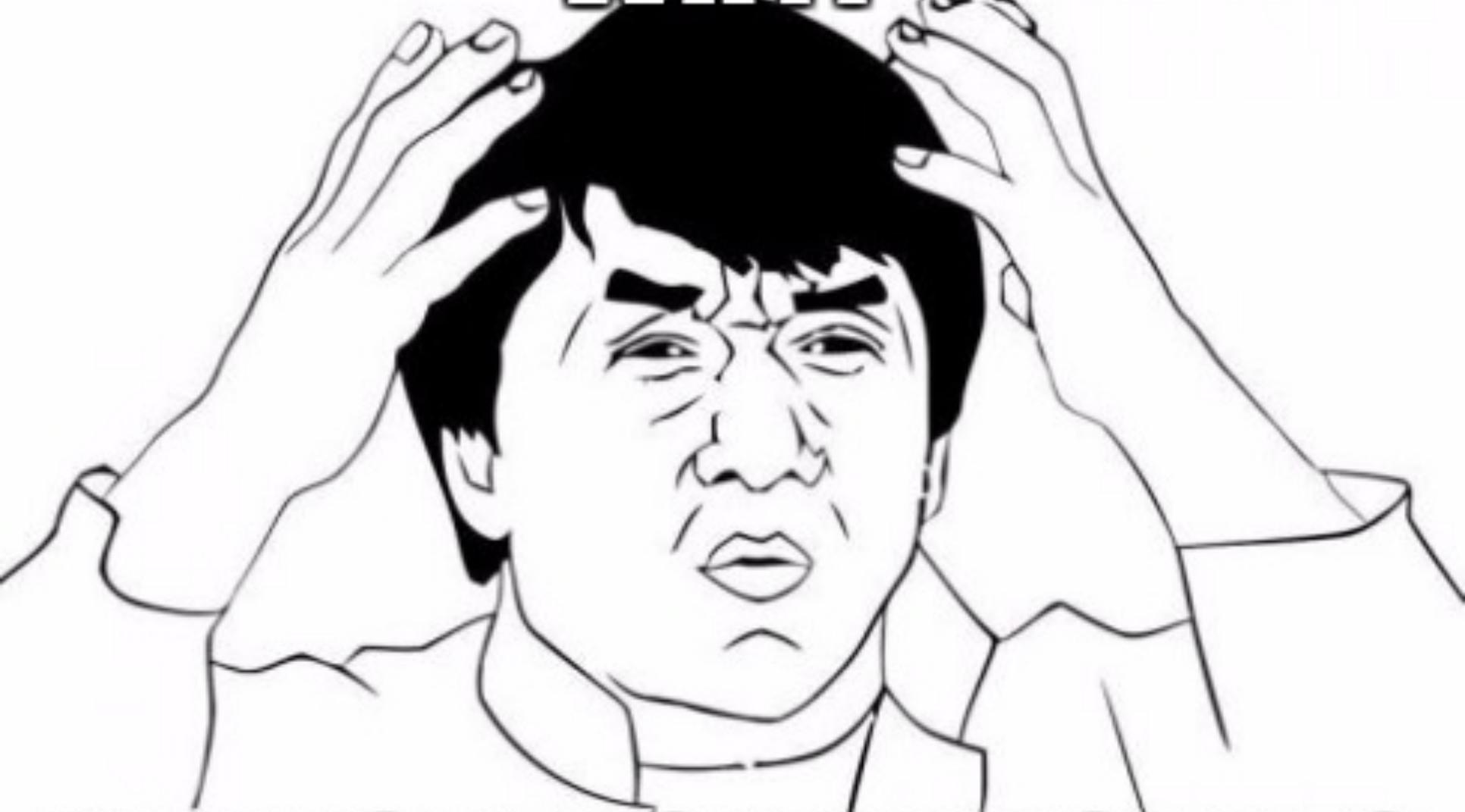
SEARCH BROWSE RECORD SOCIAL NETWORK ANALYSIS MAP FAMILY TREES LABS

Choose graph: --- Currently viewing: King Alexander II charter witnesses Locate an individual:

(Powered by Raphael Velt's [gexf-je](#) software.)



**WHY?**

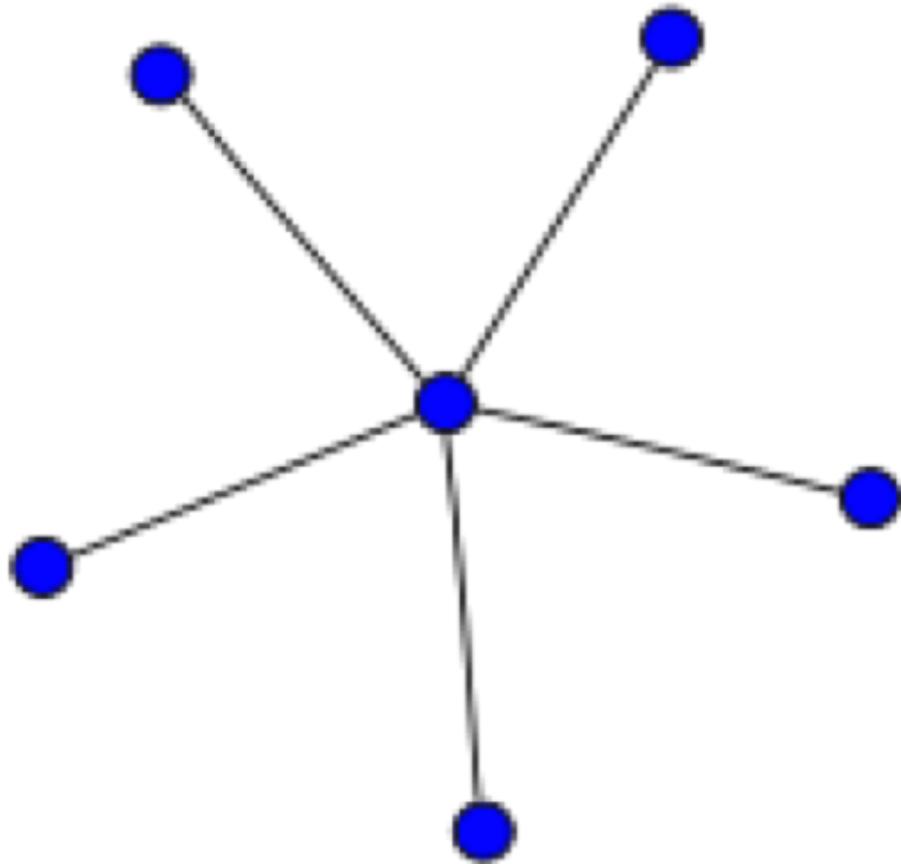


**WHY WOULD YOU EVEN DO THAT?**

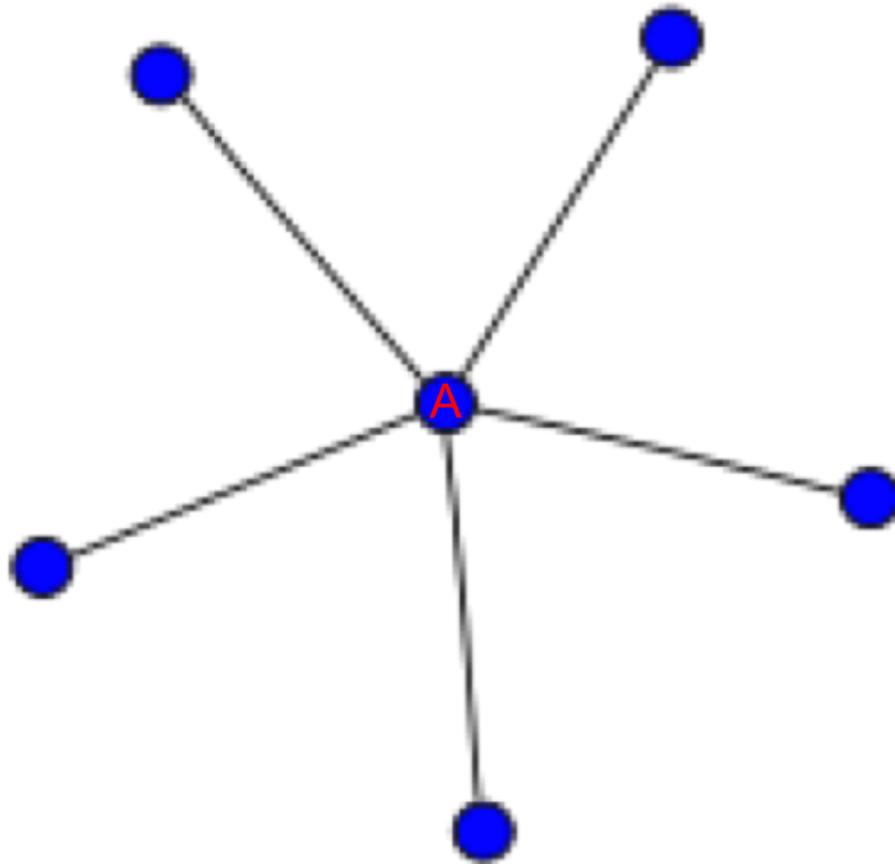
“It’s pretty (and fashionable)”

*Ondrej Tichy*

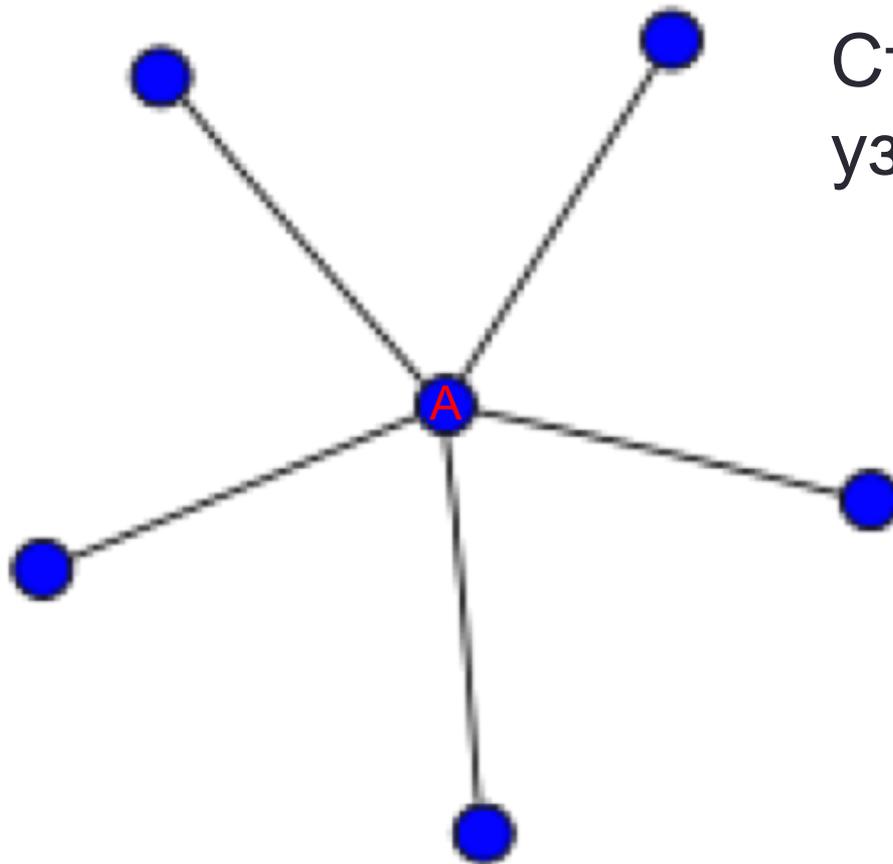
Наглядность + математическая  
строгость (#WinningCombo!)



Наглядность + математическая  
строгость (#WinningCombo!)

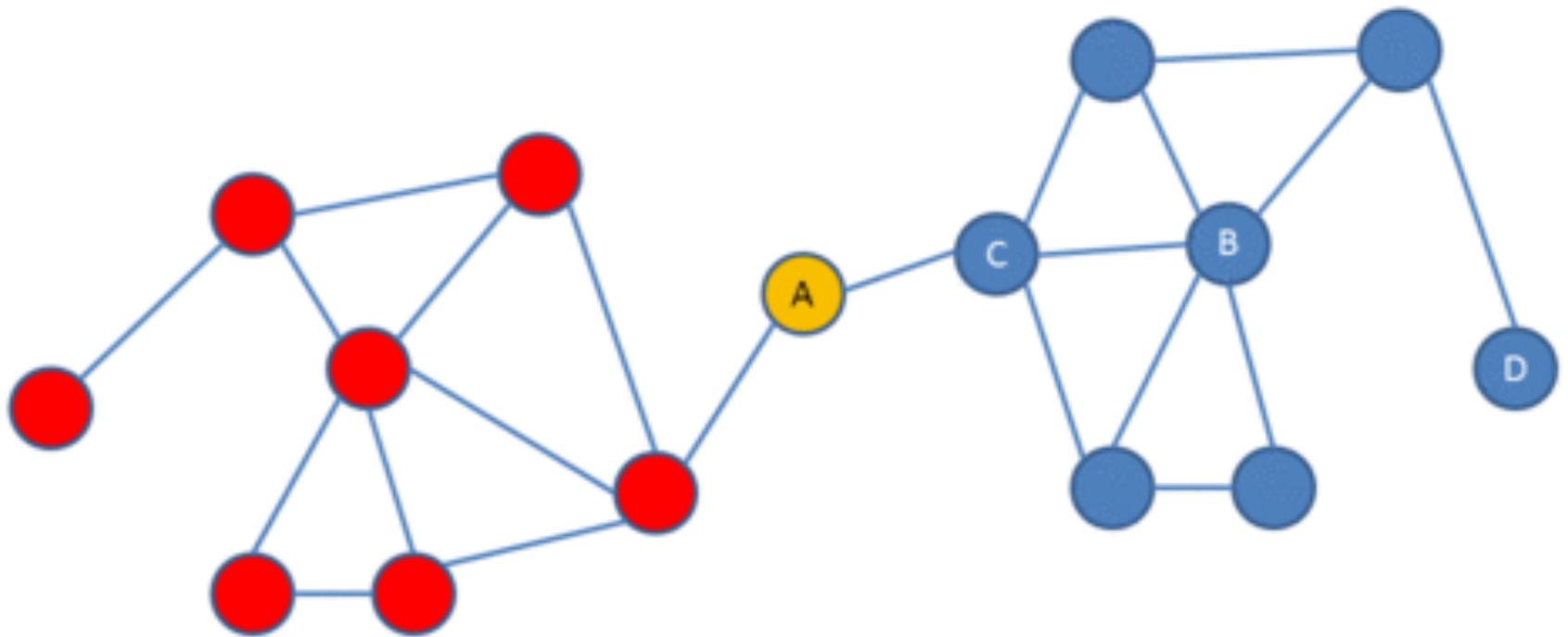


# Наглядность + математическая строгость (#WinningCombo!)



Степень (degree)  
узла A = 5

# Наглядность + математическая строгость (#WinningCombo!)



# Графы позволяют:

- Вычислять значимость отдельных узлов по разным критериям центральности
  - Степень
  - Взвешенная степень
  - Центральность по посредничеству
  - Центральность по близости
  - ...
- Анализировать пути
  - распространения информации
  - передачи болезней
  - транспортировки ресурсов
  - ....
- Выделять сообщества с опорой на данные (видели в VK)

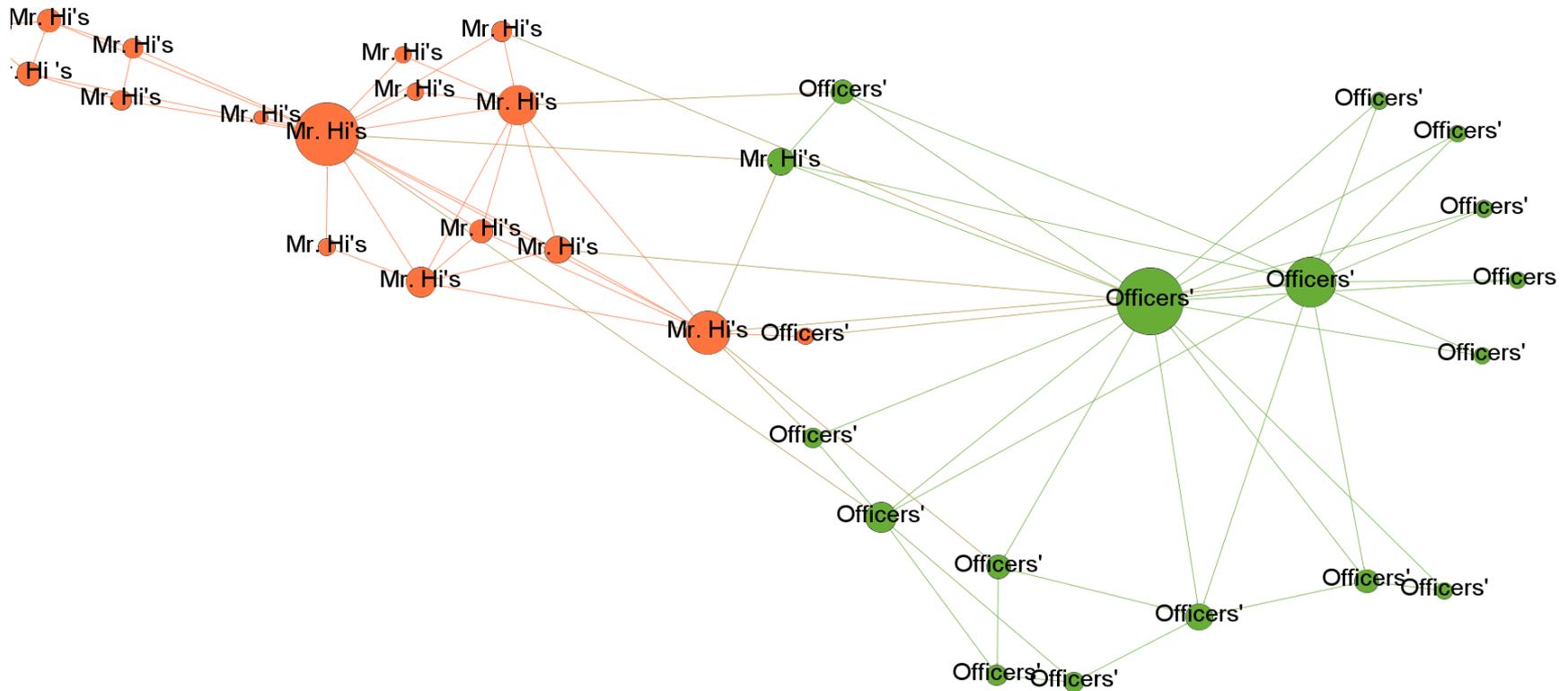
# Сообщества (тизер №2): Karate club



## Сообщества (тизер №2): Karate club

- A social network of a karate club studied by Wayne W. Zachary from 1970 to 1972
- Links capture interactions between the club members *outside the club*
- During the study a conflict arose between the administrator "John A" and instructor "Mr. Hi" (pseudonyms), which led to the split of the club into two.

# Сообщества (тизер №2): Karate club



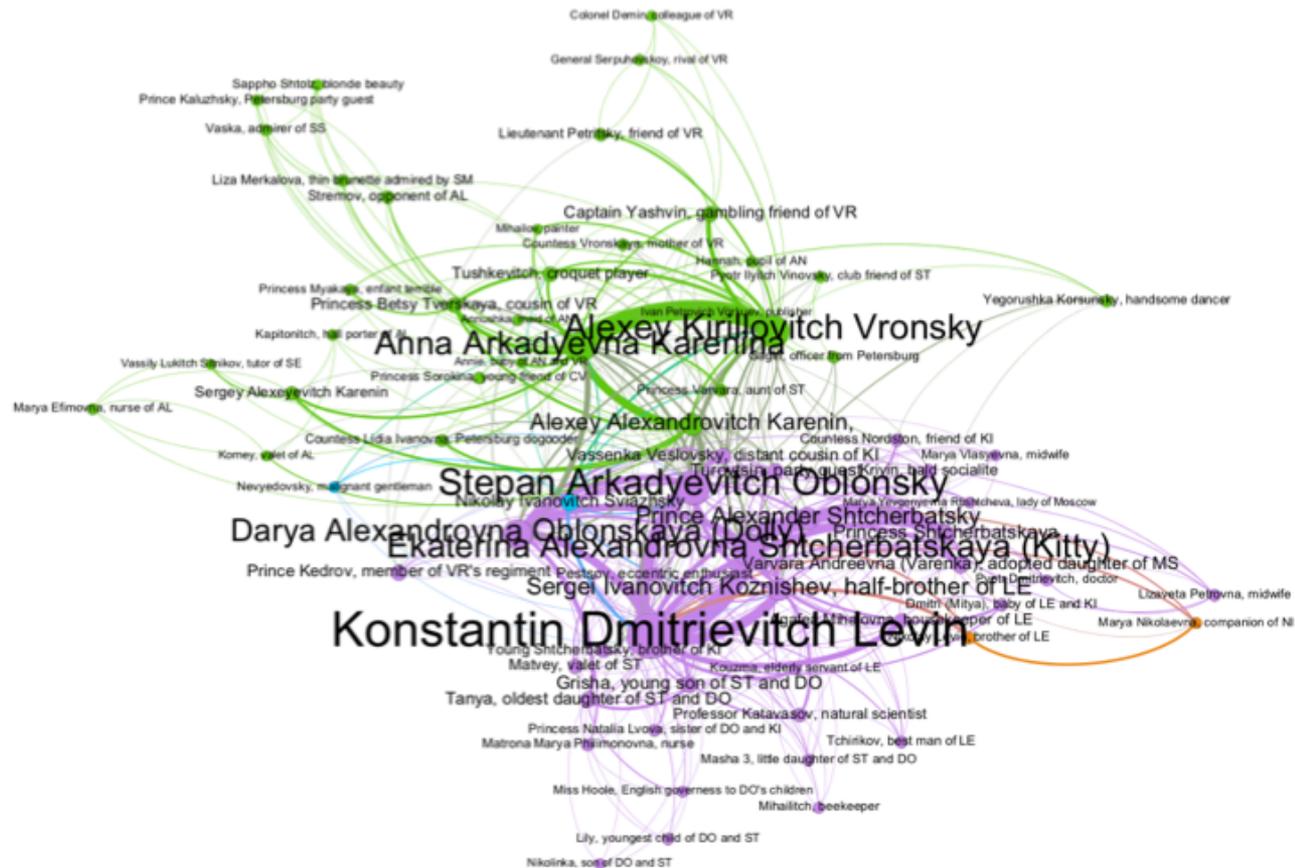
СЕТИ

В ВЫМЫШЛЕННЫХ МИРАХ

---

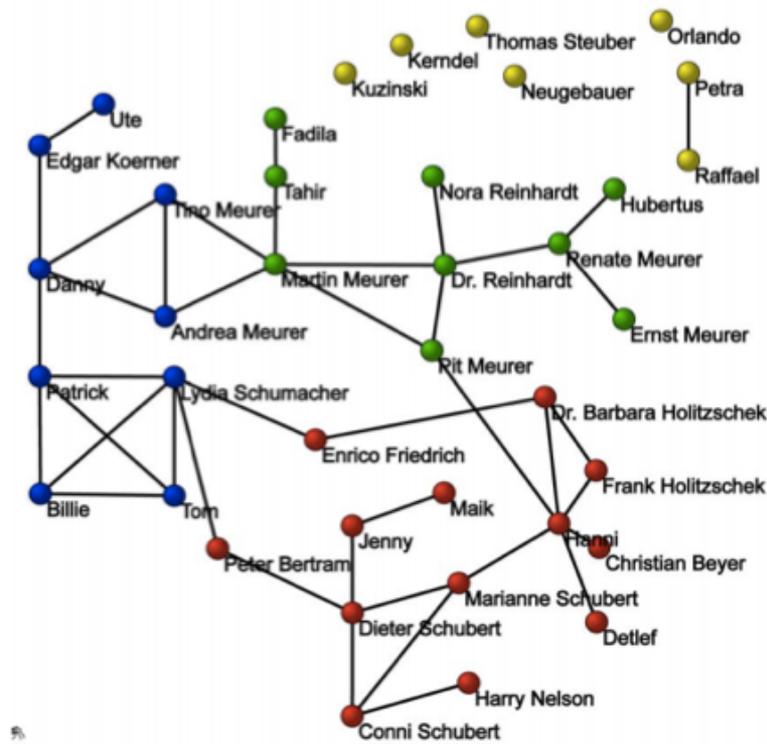
# Предыстория (origins of literary network analysis)

- Donald Knuth. Stanford GraphBase (1994) — еще нет анализа, но уже есть сети



# Предыстория

- Schweizer T., Schnegg M. Die soziale Struktur der „Simple Storys“: Eine Netzwerkanalyse. (1998)
- Антропологи решили проанализировать роман современного писателя с 38 персонажами



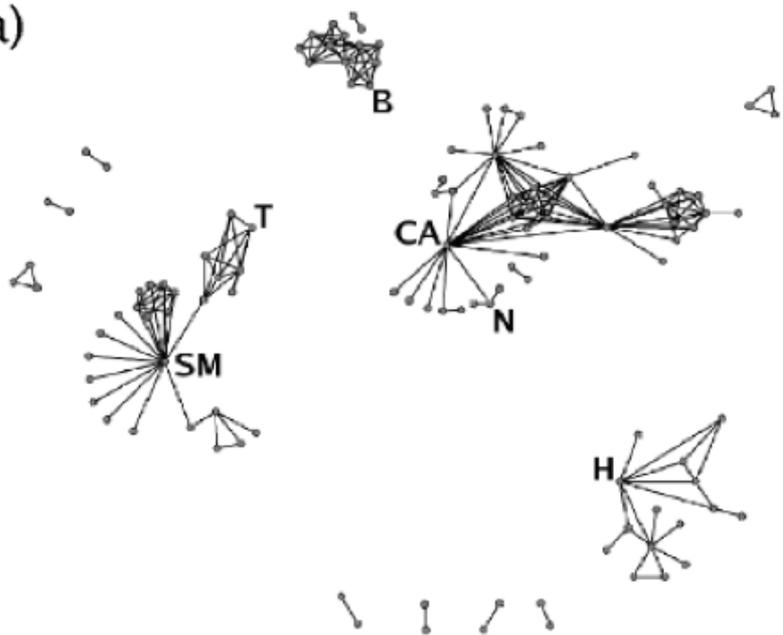
# Вселенная Marvel...



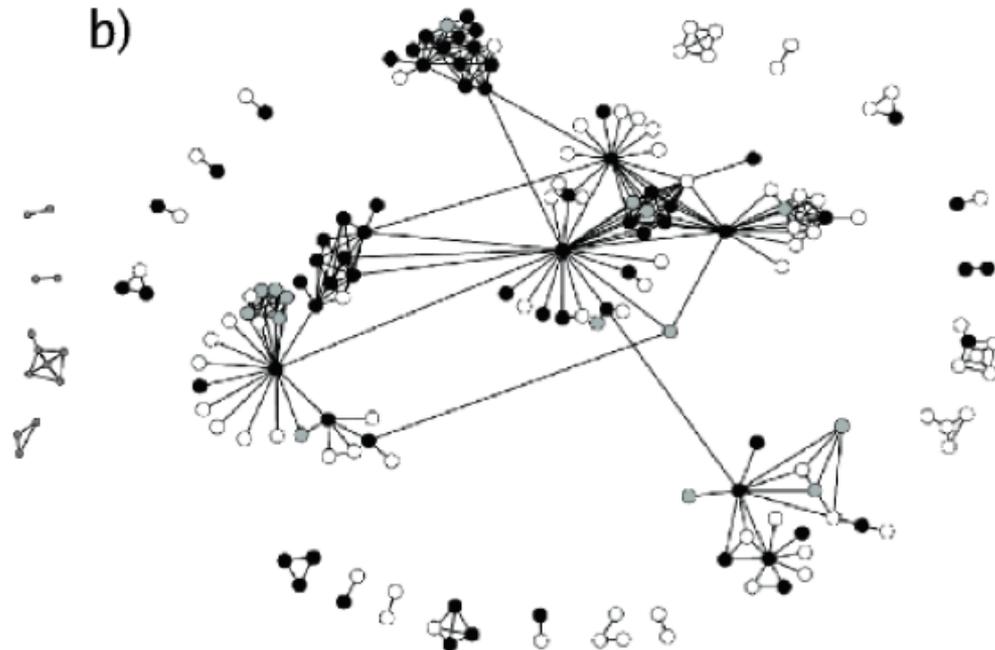
...вращается вокруг Капитана Америки!

- Alberich, R., Miro-Julia, J., Rossello, F. (2002), Marvel universe looks almost like a real social network.
- P. M. Gleiser. How to become a superhero. Journal of Statistical Mechanics: Theory and Experiment, (09):P09020, 2007.

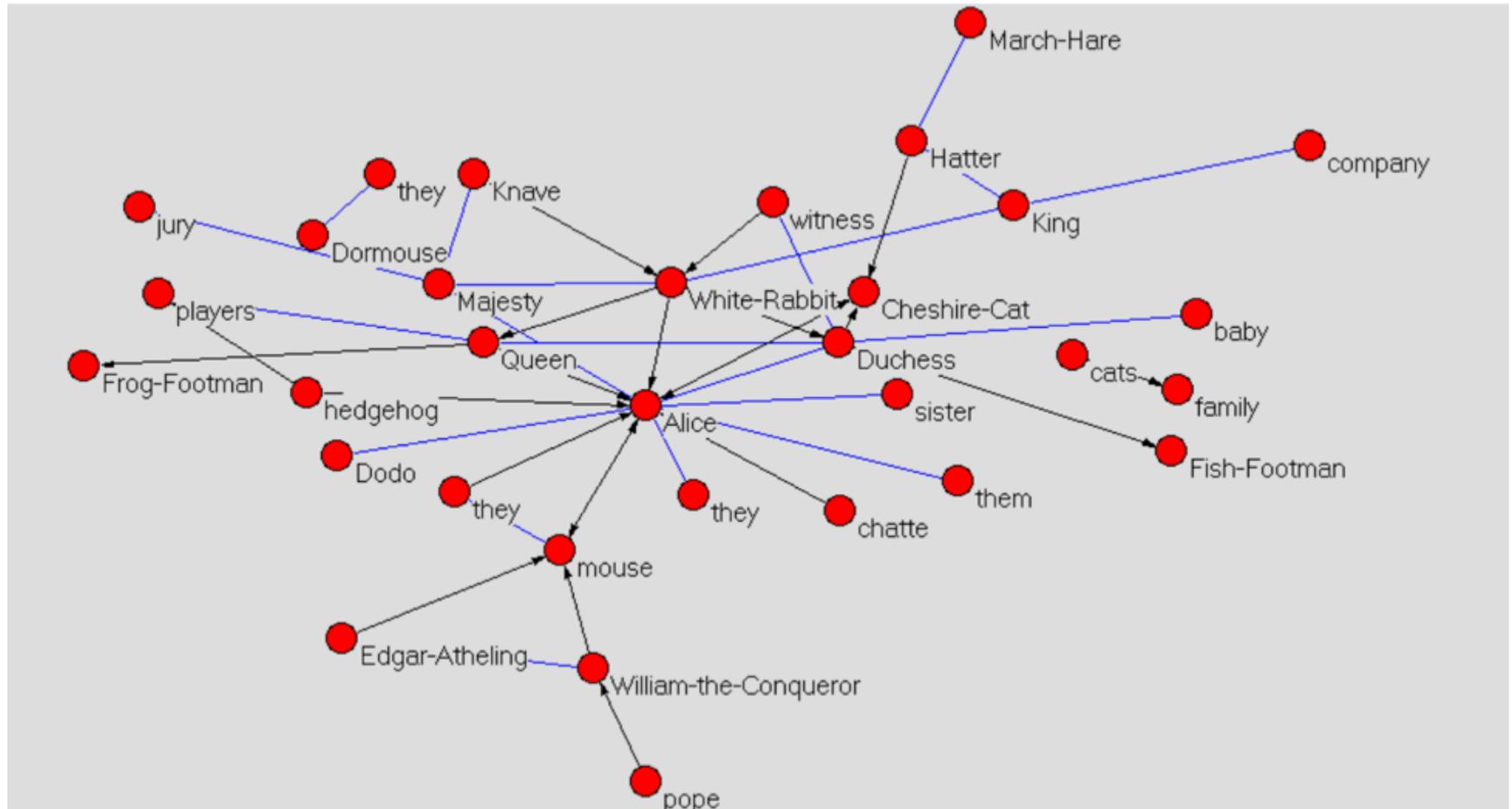
a)



b)



# Гордость и предубеждение, Алиса и кролик, Ромео и Джульетта...

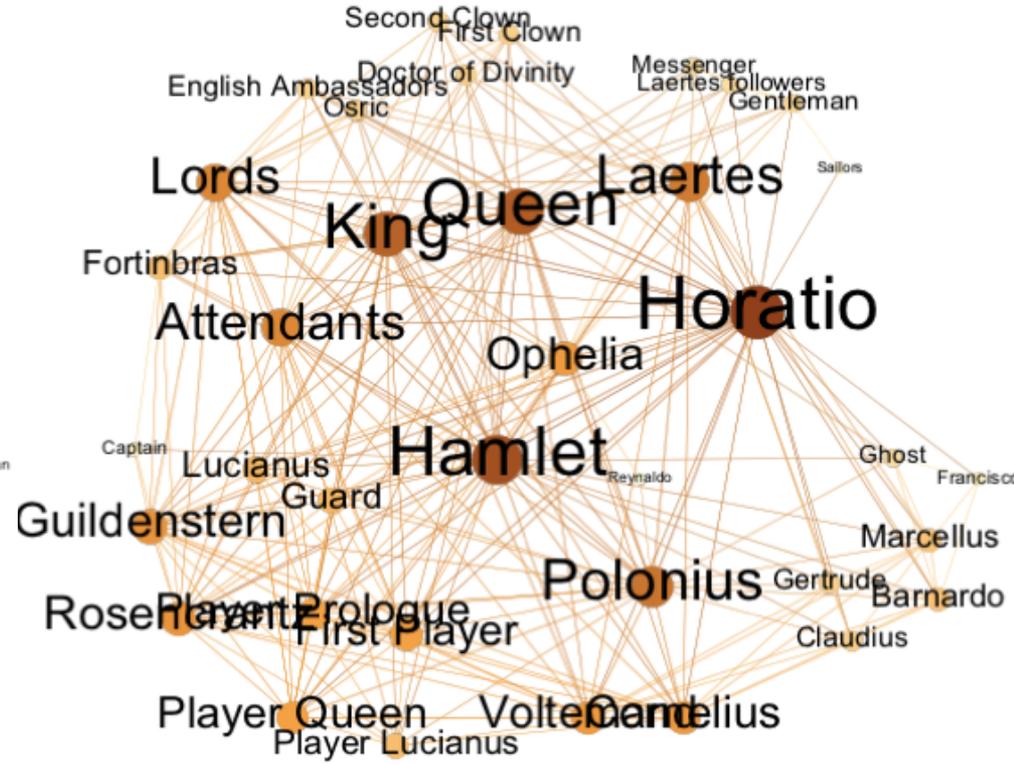
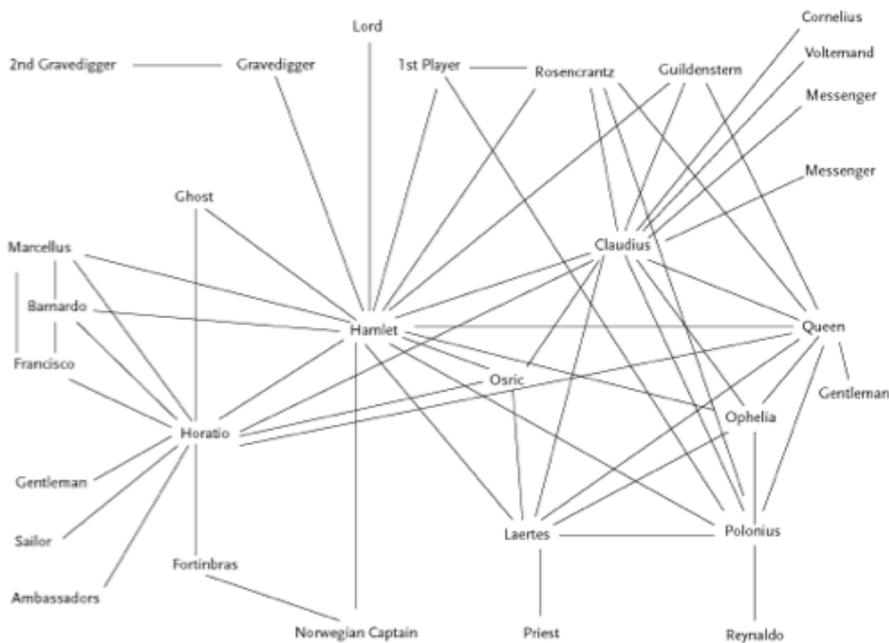


# Гордость и предубеждение, Алиса и кролик, Ромео и Джульетта...

- James Stiller, Daniel Nettle, and Robin I. M. Dunbar (2003) The Small World of Shakespeare's Plays. *Human Nature* 14(4):397---408.
- "Weak Links and Scene Cliques Within the Small World of Shakespeare," *Journal of Cultural and Evolutionary Psychology* 3, no. 1 (2005)
- Elson, D. K., Dames, N. and McKeown, K. (2010), Extracting Social Networks from Literary Fiction, *Proceedings of ACL 2010*, Uppsala, Sweden.
- J. Rydberg-Cox. Social Networks and the Language of Greek Tragedy. *Journal of the Chicago Colloquium on Digital Humanities and Computer Science*, 1(3):11, 2011.
- Agarwal A., Corvalan A., Jensen J., Rambow O. (2012), Social network analysis of Alice in Wonderland. *Proceedings of the NAACL HLT 2012 Workshop on Computational Linguistics for Literature*, pages 88–96, Montreal, Canada.
- ....

# «Переоткрытие» сетевого анализа в художественном тексте

- Франко Моретти, «Теория сетей и анализ сюжета» (2011)

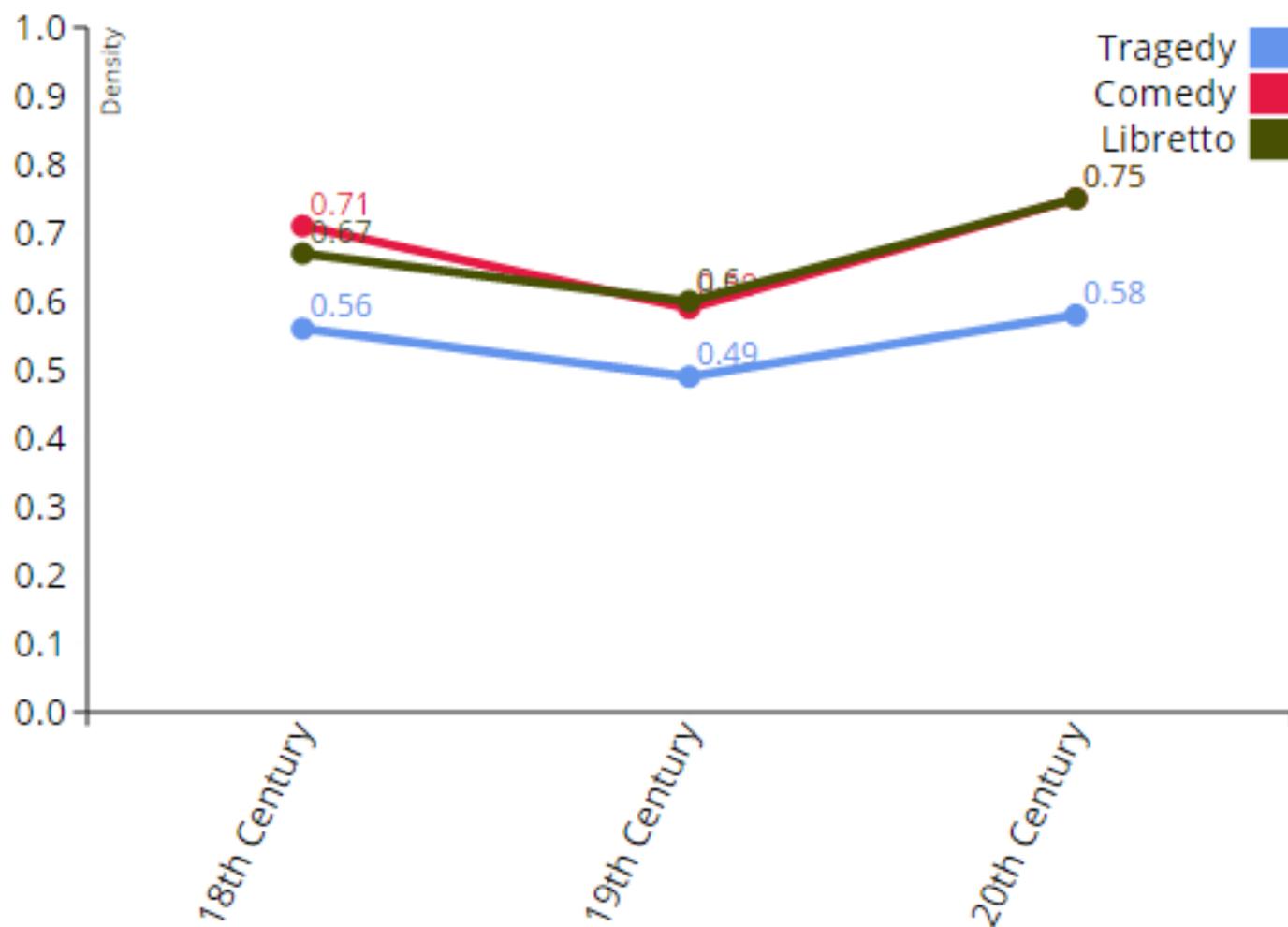


# Сетевой анализ литературы сегодня

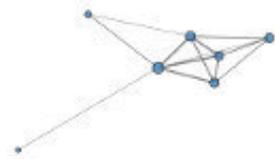
- Тренд 1: автоматизация и масштабирование
  - отдельные тексты -> сотни текстов -> тысячи текстов
  - формализация и автоматизация извлечения сетей
  - статистика и выявление общих закономерностей
- Тренд 2: анализ, интерпретация, сравнительные исследования



# Средняя плотность сети



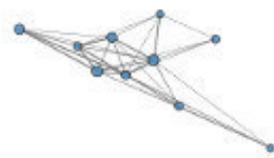
# Увеличение: Гёте начинает подражать Шекспиру:



Schlegel, C, 1746



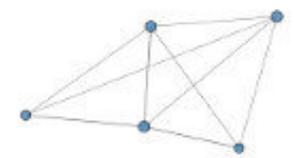
Gellert, DzS, 1747



Krüger, DCoDMzeAzg, 1747



Schlegel, DsS, 1747



Lessing, DodwF, 1747



Brawe, DF, 1758



Lessing, P, 1759



Cronegk, DM, 1760



Wieland, KvP, 1760



Gessner, EuA, 1762



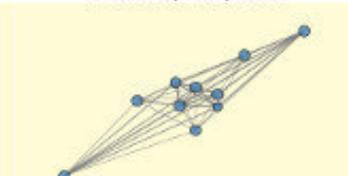
Lessing, EG, 1772



Wieland, A, 1773



Goethe, GvBmdeH, 1773



Goethe, SoDvW, 1773



Leisewitz, DP, 1774



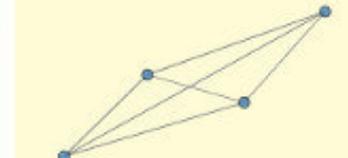
Leisewitz, JvT, 1776



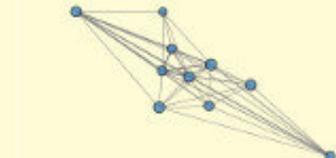
Klinger, DnA, 1776



Klinger, DZ, 1776



Goethe, DM, 1776



Goethe, S, 1776



Törring, AB, 1780



Bretzner, BuCoDEadS, 1781



Schiller, DR, 1781

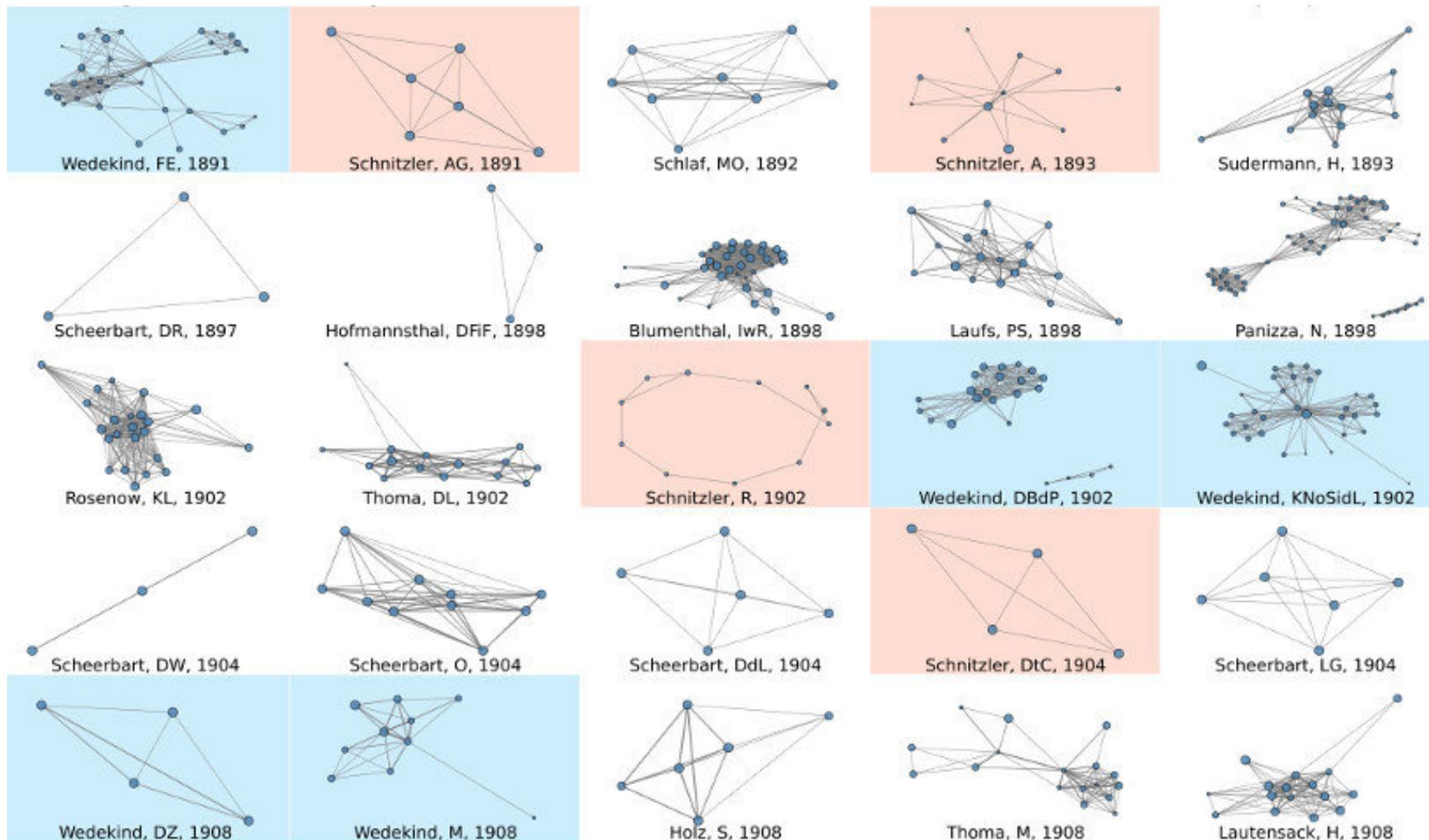


Auenbrugger, DRoDuViHaE, 1781



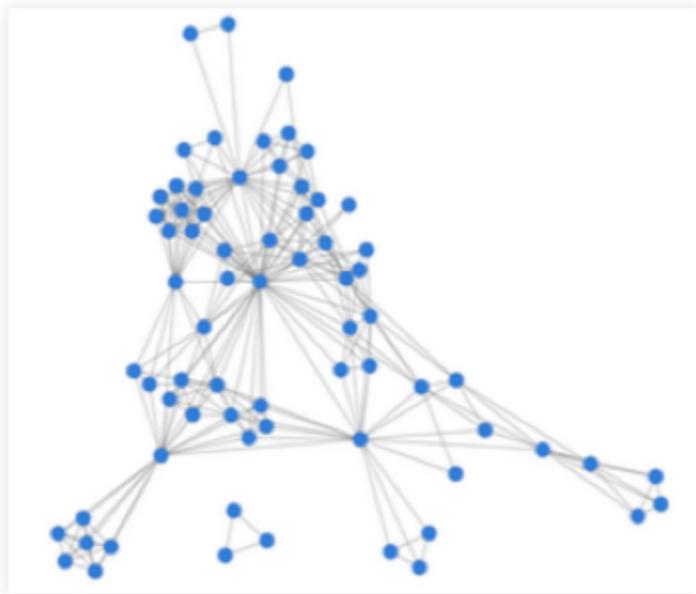
Schink, DnDF, 1782

# Увеличение: пьеса «Хоровод»



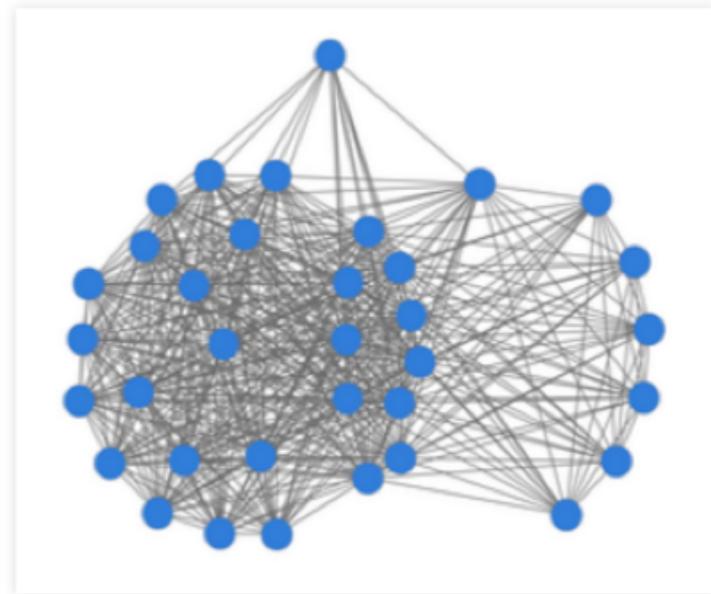
# «Коммунистическая» драма?

Goethe, *Götz* (1773)  
Drama des ›großen Individuums‹



Aristokratisches Modell?

Mühsam, *Judas* (1921)  
Drama der Masse

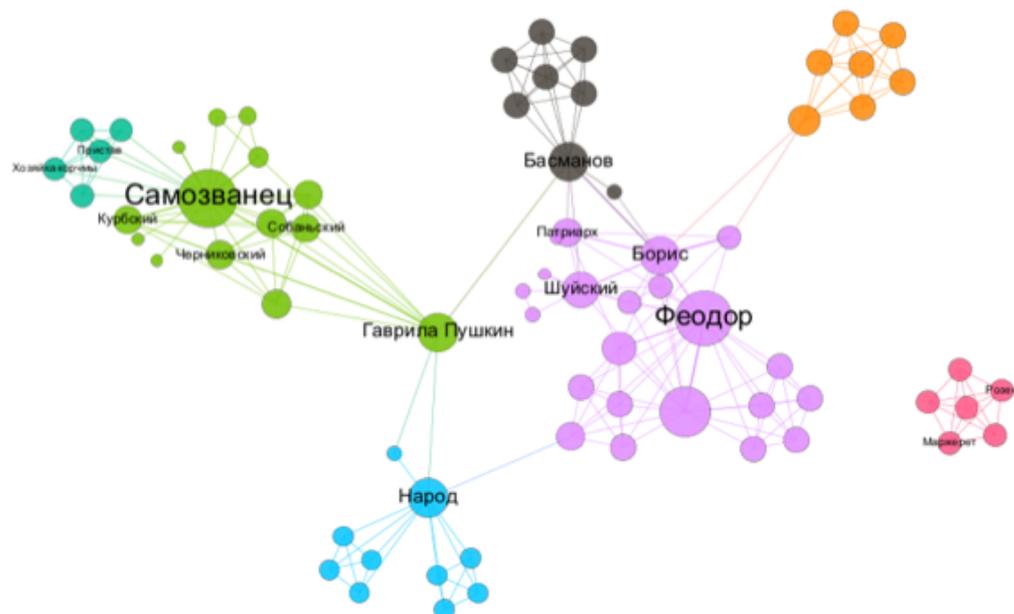
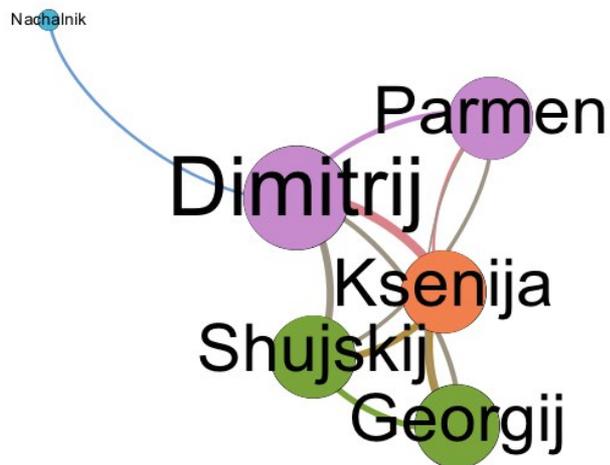


Kommunistisches Modell?

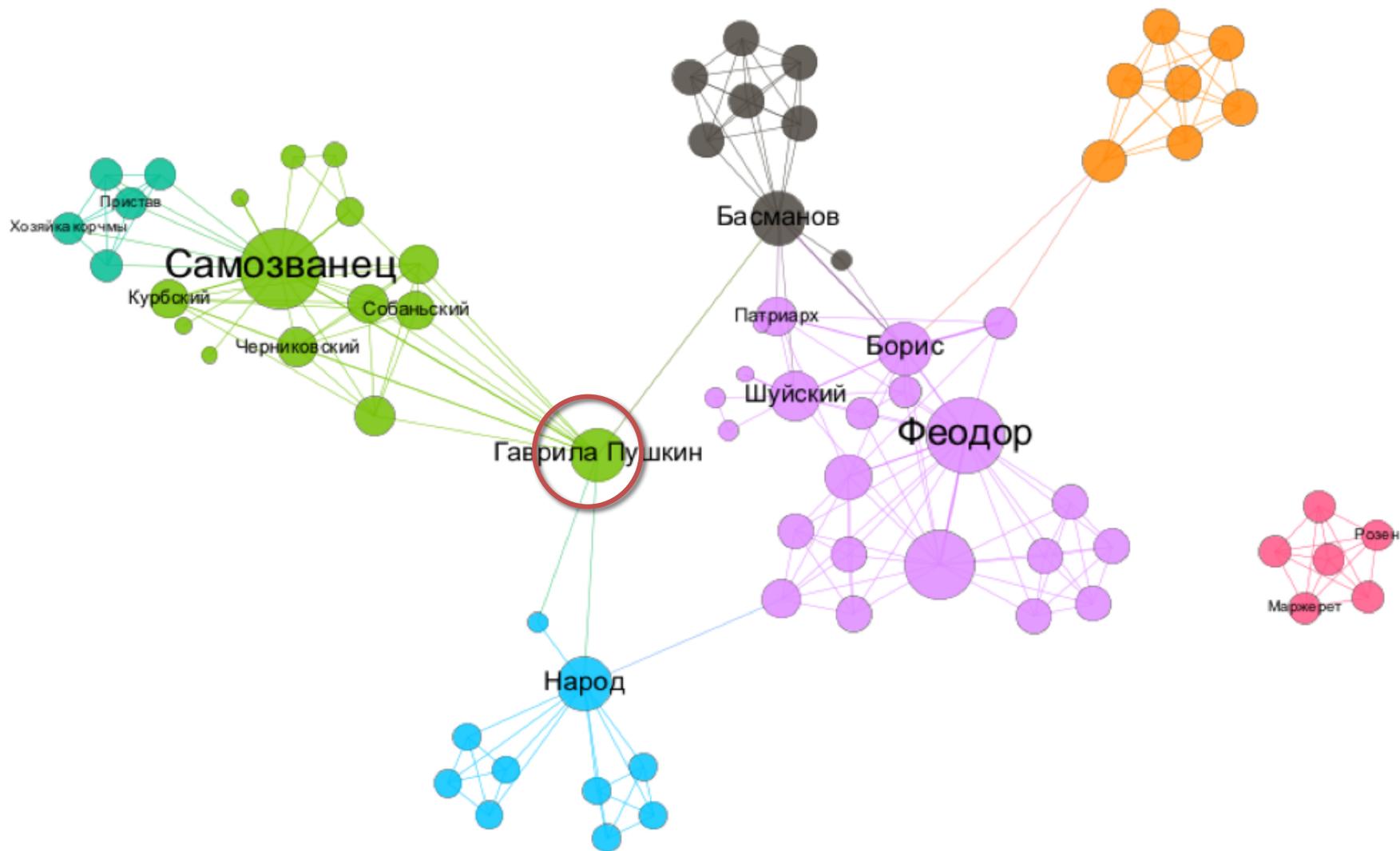
# НА РУССКОМ МАТЕРИАЛЕ

---

# Классицизм против романтизма



# Пушкин на посылках



# Гаврила Пушкин на посылках



# Корпус русской драмы с сетями

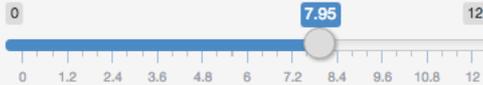
- <https://rus.dracor.org>
- 85 пьес (и мы растем!)
- Версия с красивыми сеточками:
- <https://shiny.dracor.org/>

# Russian Drama Corpus (RusDraCor): Showcase

Choose a play to visualize from a list:

Gogol\_-\_Revizor\_-\_library

Select charge:



Choose a metric for nodes size:

Degree

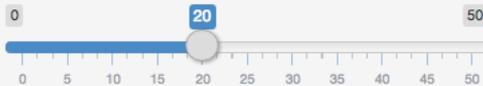
Nodes size:



Choose clusterization algorithm:

cluster\_edge\_betweenness

Select font size:



Edges size:



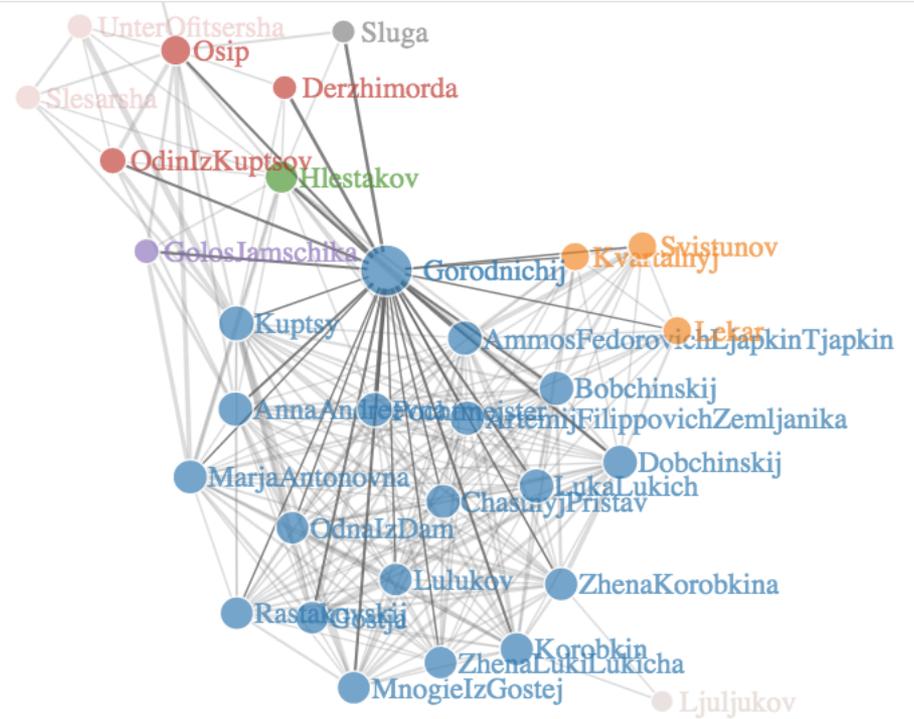
Graph

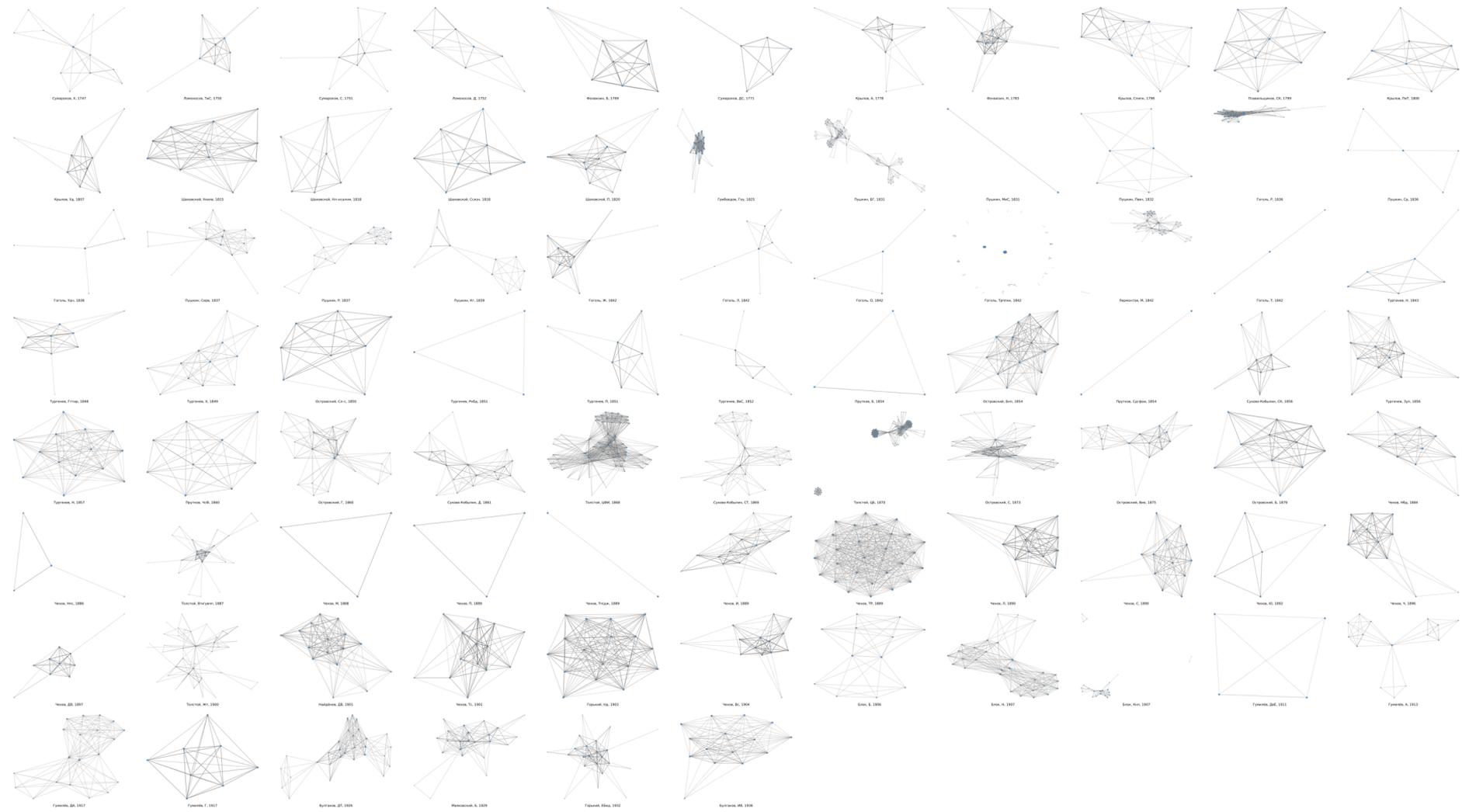
Edges

Vertices

Weights matrix

About





Савицкий, А. 1947

Александр, Т.С. 1950

Сидоров, С. 1951

Александр, Д. 1952

Александр, Б. 1959

Сидоров, Д.С. 1973

Кликин, А. 1978

Александр, В. 1983

Кликин, С.В. 1988

Полосинский, Д. 1993

Кликин, Д.В. 2000

Кликин, А. 1987

Сидоров, Н.В. 1985

Сидоров, И.И. 1988

Сидоров, С.В. 1989

Сидоров, П. 1990

Гурвич, Ф. 1925

Пурин, В. 1931

Пурин, М.С. 1933

Пурин, С. 1935

Фарш, П. 1936

Пурин, С. 1938

Фарш, П. 1938

Пурин, С. 1937

Пурин, А. 1937

Пурин, В. 1939

Фарш, М. 1942

Фарш, П. 1942

Фарш, Д. 1942

Фарш, Т.В. 1942

Радостин, А. 1942

Фарш, Т. 1942

Турин, Н. 1943

Турин, П.И. 1944

Турин, А. 1944

Орловский, С.С. 1950

Турин, П.И. 1951

Турин, П. 1951

Турин, М.С. 1952

Пурин, А. 1954

Орловский, С. 1954

Пурин, С.В. 1954

Сидоров, Д.С. 1956

Турин, П. 1956

Турин, А. 1957

Пурин, М.В. 1960

Орловский, Т. 1968

Сидоров, Б. 1961

Турин, П.И. 1968

Сидоров, С.Т. 1969

Турин, А. 1973

Орловский, С. 1973

Орловский, С. 1975

Орловский, С. 1976

Турин, М. 1984

Турин, П. 1988

Турин, В.В. 1987

Турин, П. 1988

Турин, П. 1989

Турин, С. 1990

Турин, А. 1992

Турин, А. 1990

Турин, А. 1987

Турин, М. 1990

Турин, А. 1991

Турин, Т. 1991

Турин, П. 1993

Турин, С. 1994

Турин, А. 1996

Турин, А. 1997

Турин, М. 1997

Турин, Д. 1991

Турин, А. 1993

Турин, Д. 1997

Турин, П. 1997

Турин, А. 1998

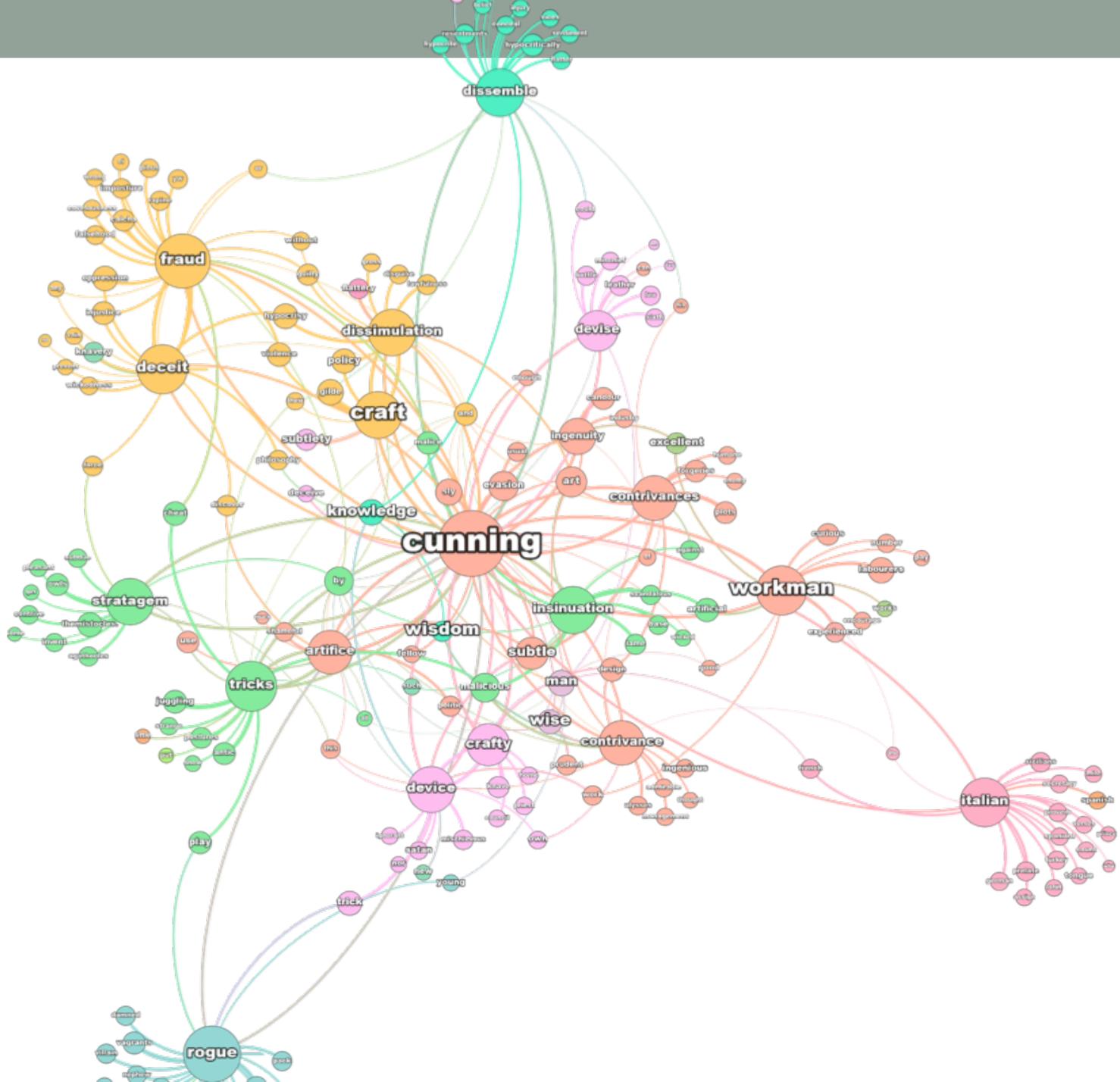
Турин, А. 1999

Турин, А. 1992

Турин, А. 1996

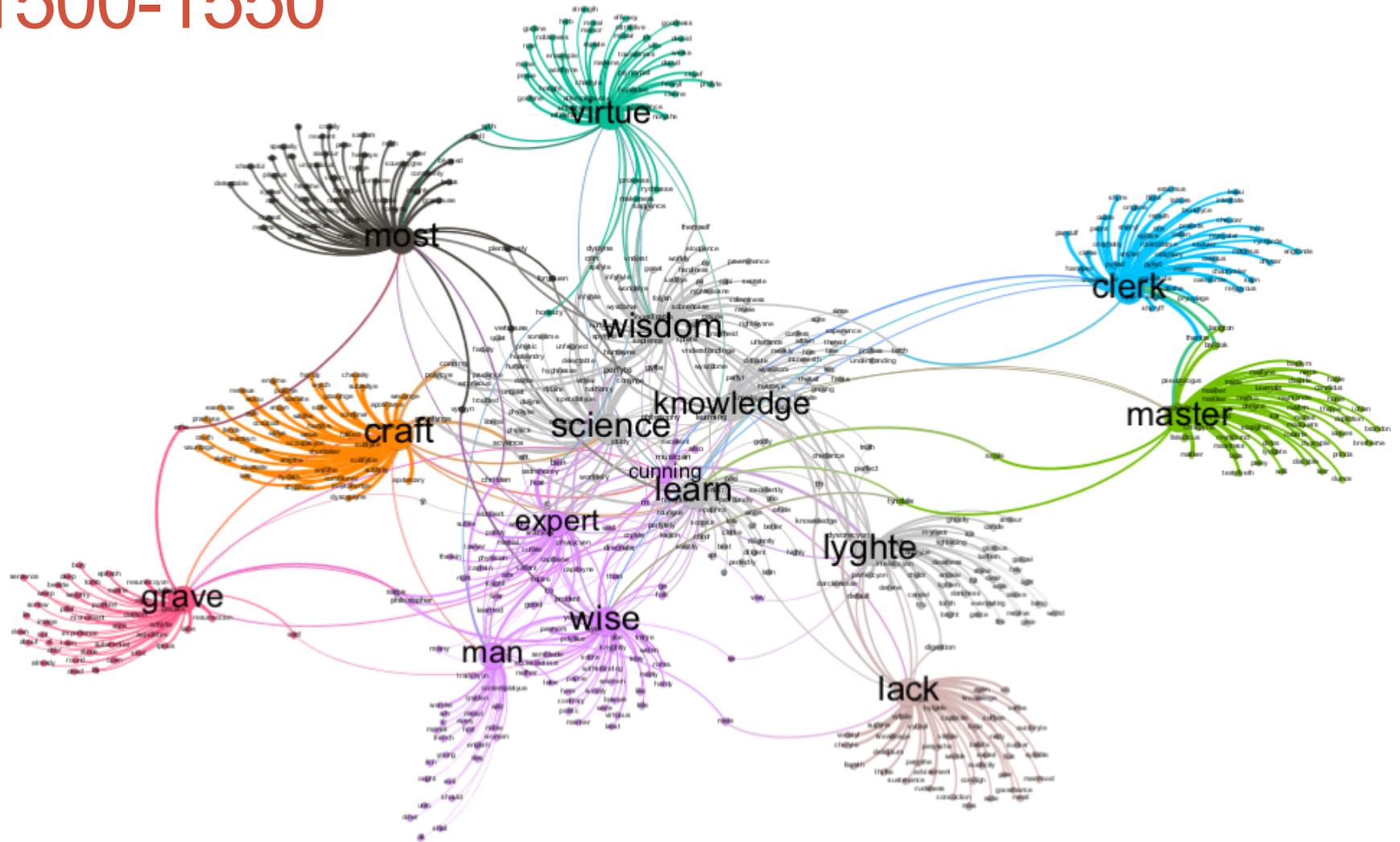
# СЕТИ В ЛИНГВИСТИКЕ

---



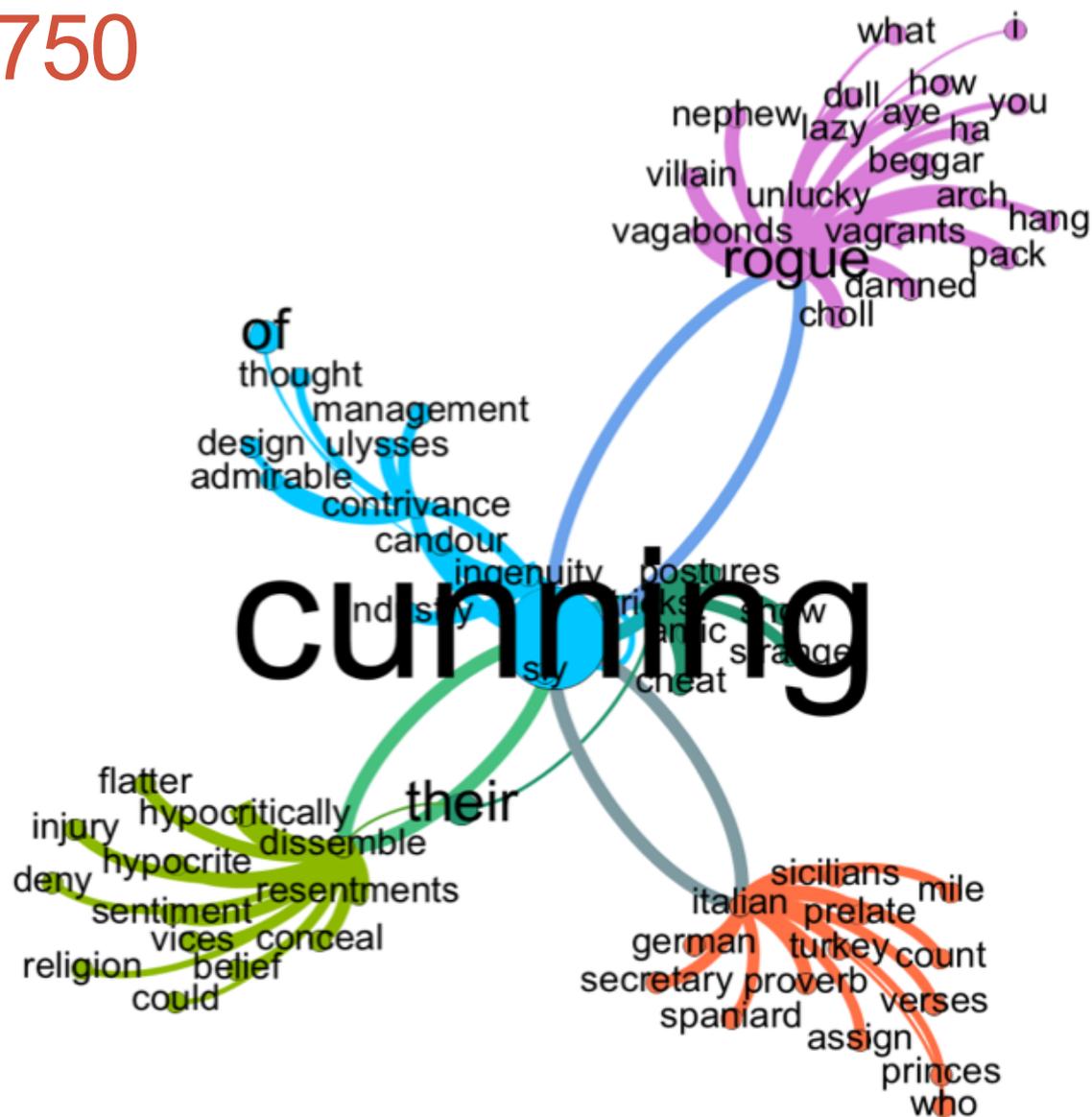
# Коллокации слова 'cunning'

## 1500-1550



# Коллокации слова 'cunning'

## 1700-1750



GEPHI

---

# Gephi

- Простой инструмент для анализа сетей (не надо уметь программировать)
- "like Photoshop but for graph data" ([gephi.org](http://gephi.org))
- Открытый и бесплатный
- Написан на Java и работает на всех основных ОС
- Расширяемый (можно создавать плагины)
- ...немного глючный 😊

# Тutorials

- <https://gephi.org/users/>

# Gerhi: простейшая демонстрация

- Сначала сделаем свой игрушечный граф

# А как засунуть/вытащить готовую сеть в/из Gephi?

Gephi: форматы данных

- CSV (можно сделать таблицу в Excel/OpenOffice)
- GEXF
- GraphML,
- GML
- Pajek
- ...еще миллион

# GEXF

```
<gexf>
  <graph defaultedgetype="undirected" mode="static">
    <nodes>
      <node id="1" label="A"/>
      <node id="2" label="B"/>
      <node id="3" label="C"/>
      <node id="4" label="D"/>
      <node id="5" label="E"/>
    </nodes>
    <edges>
      <edge id="0" source="1" target="2"/>
      <edge id="1" source="1" target="3"/>
      <edge id="2" source="2" target="4"/>
      <edge id="3" source="3" target="4"/>
      <edge id="4" source="4" target="5"/>
      <edge id="6" source="1" target="1"/>
    </edges>
  </graph>
</gexf>
```

# Практика 1

1. Открываем папку  
<https://goo.gl/ShjAKn>
2. Сохраняем себе файл  
**1\_scotland.gexf** (Дарители и получатели даров в средневековой Шотландии )
3. Открываем Gephi
4. **Файл->Открыть-> 1\_scotland.gexf**

# CSV

- **Comma-Separated Values** (значения, разделенные запятой)
- Source, Type, Target, Weight
- Fedor\_Ivanovich\_Dolokhov, Undirected, Vasily\_Denisov, 10
- Fedor\_Ivanovich\_Dolokhov, Undirected, Petya\_Rostov, 12
- AndreyBolkonsky, Undirected, Boris\_Drubetskoy, 3
- Anna\_Pavlovna\_Scherer, Undirected, Bilibin, 1
- Nikolai\_Rostov, Undirected, Sonya\_Rostova, 35
- Count\_Rostopchin, Undirected, Vereshchagin, 1
- Anna\_Pavlovna\_Scherer, Undirected, Pierre\_Bezukhov, 12
- Anatole\_Kuragin, Undirected, Vasili\_Kuragin, 1
- Anna\_Pavlovna\_Scherer, Undirected, Vasili\_Kuragin, 4

## Практика 2

1. Открываем папку  
<https://goo.gl/ShjAKn>
2. Сохраняем себе файл  
**StormOfSwords.csv**
3. Открываем Gephi
4. **Файл->Import Spreadsheet->  
StormOfSwords.csv**

# Укладка графа (layout)

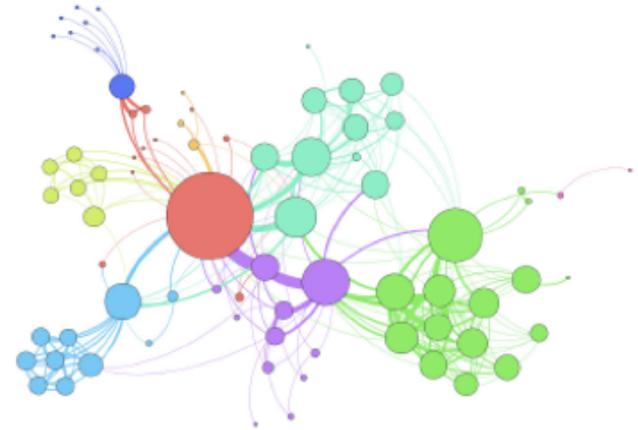
- ...это отдельная наука на стыке математики, computer science и физики
- Может подчеркивать и выделять разные свойства графа

# Укладка графа (layout)

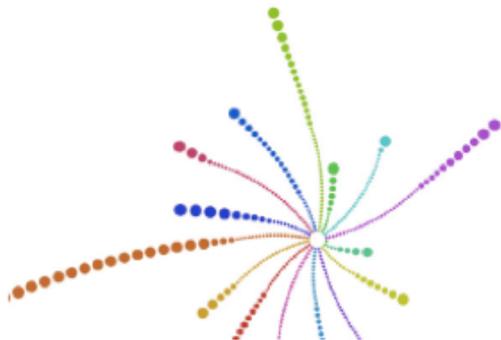
*OpenOrd*



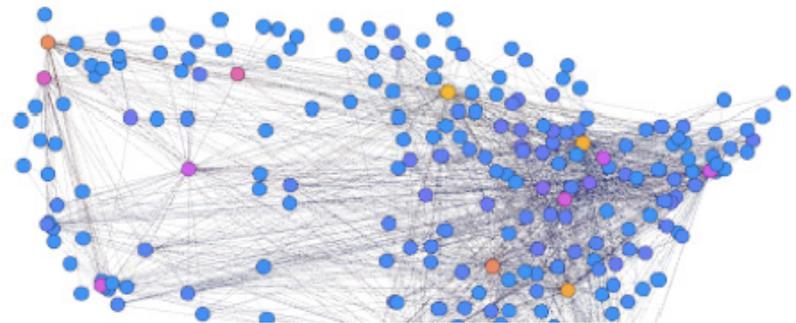
*ForceAtlas 2*



*Radial Axis*



*GeoLayout*



## Next time

- Выделение сообществ
- Где брать сети
- Как делать сети самому (не в гефи)

# Полезные ссылки

- [https://en.wikipedia.org/wiki/Social\\_network\\_analysis](https://en.wikipedia.org/wiki/Social_network_analysis)
- <http://www.leonidzhukov.net/hse/2014/socialnetworks/>
- <https://gephi.org/users/>
- <http://www.martingrandjean.ch/gephi-introduction/>
- <https://github.com/gephi/gephi/wiki/Datasets>
- <http://konect.uni-koblenz.de/networks/>
- <http://www-cs-faculty.stanford.edu/~knuth/sgb.html>